LANGUAGE & LITERATURE

R406 N21

PUBLIC LIBRARY
APR 1
DETROIT

SPEECH MONOGRAPHS

Published by THE SPEECH ASSOCIATION OF AMERICA

VOLUME XXIV MARCH, 1957 No. 1

Greek and English Consonants — JAMES W. ABEL

Some Determinants of the Effect of Oral Communication in Producing Attitude Change and Learning — DAVID K. BERLO and HALBERT E. GULLEY

A Study of the Characteristics of an Audience Viewing a Specific Telecourse — JOHN R. SHEPHERD

An Experimental Study of Retention In Educational Television — DANIEL W. MULLIN

An Experimental Study of the Acquistion of Information from Three Types of Recorded Television Presentations — JOHN HOLWAY ULRICH

Attitudes of Speech Defectives Toward Humor Based on Speech Defects — LEOLA SCHAPER HOROWITZ

A Comparison of the Chain Associations of Nursery School and Kindergarten Children to Action-Picture Stimuli — MARIAN NELSON FERGUSON

The Construction and Testing of a Forced Choice Scale for Measuring Speaking Achievement — KEITH BROOKS

The Effects of a Language Training Program on Foreign Soundingness — DALE J. LUNDEEN, PAUL H. PTACEK, CLARK D. STARR, ERNEST H. HENRIKSON

R406 N21
v.24 1957 c.1

SPEECH MONOGRAPHS

PUBLISHED BY
THE SPEECH ASSOCIATION OF AMERICA

WALDO W. BRADEN, *Executive Secretary*
Louisiana State University, Baton Rouge 3, Louisiana

EDITOR

HOWARD GILKINSON
309 Folwell Hall
University of Minnesota
Minneapolis 14, Minnesota

EDITORIAL BOARD

JOHN W. BLACK, *The Ohio State University*
WARREN GUTHRIE, *Western Reserve University*
MARIE K. HOCHMUTH, *University of Illinois*
JACK MATTHEWS, *University of Pittsburgh*
PAUL MOORE, *Northwestern University*
ORVILLE L. PENCE, *University of Washington*
DAVID W. THOMPSON, *University of Minnesota*
EDGAR E. WILLIS, *University of Michigan*

CONTRIBUTING EDITORS

J. JEFFERY AUER, *University of Virginia*
JAMES W. CLEARY, *University of Wisconsin*
CLYDE W. DOW, *Michigan State University*
FRANKLIN H. KNOWER, *The Ohio State University*

* * *

Speech Monographs is published in the months of March, June, August, and November by the Speech Association of America, Louisiana State University, Baton Rouge 3, Louisiana. Entered as second-class matter at the post office, Columbia, Missouri, under the Act of March 3, 1879.

Correspondence concerning issues should be addressed to the Baton Rouge office. Correspondence concerning manuscripts should be addressed to the Editor. Copyright, 1957, by the Speech Association of America.

Annual Subscription, $6.00. Single copies, $2.00.

LLP

SPEECH ASSOCIATION *of* AMERICA

Founded, 1914
Incorporated, 1950

OFFICERS

LOREN REID, *President*	University of Missouri
KENNETH G. HANCE, *Executive Vice-President*	Michigan State University
ELISE HAHN, *First Vice-President*	University of California at Los Angeles
JOHN E. DIETRICH, *Second Vice-President*	The Ohio State University
DONALD C. BRYANT, *Editor, The Quarterly Journal of Speech*	Washington University
HOWARD GILKINSON, *Editor, Speech Monographs*	University of Minnesota
HENRY L. MUELLER, *Editor, The Speech Teacher*	University of Illinois
WALDO W. BRADEN, *Executive Secretary*	Louisiana State University

OFFICERS-ELECT

J. JEFFERY AUER, *Executive Vice-President-Elect*	University of Virginia
OWEN M. PETERSON, *Executive Secretary-Elect*	Louisiana State University
KARL F. ROBINSON, *Editor-Elect, The Speech Teacher*	Northwestern University

ADMINISTRATIVE COUNCIL

Term Expires 1957

PAUL BAGWELL, Executive Vice-President, 1952-1954; Ex Officio, 1955-1957	Michigan State University
MILTON DICKENS, Elected by SAA, 1954	University of Southern California
DALLAS C. DICKEY, Editor, *The Speech Teacher*, 1952-1954; Ex Officio 1955-1957	University of Florida
JOHN E. DIETRICH, Second Vice-President	The Ohio State University
WILLARD J. FRIEDERICH, Elected by SAA, 1954	Marietta College
ELISE HAHN, First Vice-President	University of California at Los Angeles
ORVILLE A. HITCHCOCK, Executive Secretary of SAA, 1951-1954; Ex Officio, July, 1954-June, 1957; Chairman of Finance Committee	State University of Iowa
CHARLES A. MCGLON, Elected by SAA, 1954	Southern Baptist Theological Seminary
EDYTH RENSHAW, Elected by SAA, 1954	Southern Methodist University
KARL R. WALLACE, President of SAA, 1954; Ex Officio, 1955-1957; Member of Finance Committee	University of Illinois

Term Expires 1958

BARNARD HEWITT, Elected by SAA, 1955 — University of Illinois
T. EARLE JOHNSON, Elected by SAA, 1955 — University of Alabama
EVELYN KONIGSBERG, Elected by SAA, 1955 — Washington Irving High School, New York City
THOMAS A. ROUSSE, President of SAA, 1955; Ex Officio, 1956-1958 — University of Texas
CLARENCE T. SIMON, Elected by SAA, 1955 — Northwestern University

Term Expires 1959

J. JEFFERY AUER, Editor, *Speech Monographs,* 1954-1956; Ex Officio, 1957-1959 — University of Virginia
ROBERT D. CLARK, Elected by SAA, 1956 — University of Oregon
WILBUR SAMUEL HOWELL, Editor, *The Quarterly Journal of Speech,* 1954-1956; Ex Officio, 1957-1959 — Princeton University
KARL F. ROBINSON, Elected by SAA, 1956 — Northwestern University
LESTER THONSSEN, President of SAA, 1956; Ex Officio, 1957-1959 — College of the City of New York

Term Expires 1960

WALDO W. BRADEN, Executive Secretary of SAA, 1954-1957; Ex Officio July, 1956-June, 1960 — Louisiana State University
KENNETH G. HANCE, Executive Vice-President; Ex Officio, 1958-1960 — Michigan State University
HENRY L. MUELLER, Editor, *The Speech Teacher,* 1955-1957; Ex Officio, 1958-1960 — University of Illinois
OWEN M. PETERSON, Executive Secretary of SAA, July 1957-June 1960 — Louisiana State University
LOREN D. REID, President of SAA, 1957; Ex Officio, 1958-1960 — University of Missouri

Term Expires 1962

DONALD C. BRYANT, Editor, *The Quarterly Journal of Speech,* 1957-1959; Ex Officio, 1960-1962 — Washington University
HOWARD GILKINSON, Editor, *The Speech Monographs,* 1957-1959; Ex Officio, 1960-1962 — University of Minnesota

SPEECH MONOGRAPHS

Published by THE SPEECH ASSOCIATION OF AMERICA

VOLUME XXIV MARCH, 1957 No. 1

Greek and English Consonants 1
JAMES W. ABEL

Some Determinants of the Effect of Oral Communication in Producing Attitude Change and Learning 10
DAVID K. BERLO and HALBERT E. GULLEY

A Study of the Characteristics of an Audience Viewing a Specific Telecourse 21
JOHN R. SHEPHERD

An Experimental Study of Retention In Educational Television 31
DANIEL W. MULLIN

An Experimental Study of the Acquisition of Information from Three Types of Recorded Television Presentations 39
JOHN HOLWAY ULRICH

Attitudes of Speech Defectives Toward Humor Based on Speech Defects 46
LEOLA SCHAPER HOROWITZ

A Comparison of the Chain Associations of Nursery School and Kindergarten Children to Action-Picture Stimuli 56
MARIAN NELSON FERGUSON

The Construction and Testing of a Forced Choice Scale for Measuring Speaking Achievement 65
KEITH BROOKS

The Effects of a Language Training Program on Foreign Soundingness 74
DALE J. LUNDEEN, PAUL H. PTACEK, CLARK D. STARR, ERNEST H. HENRIKSON

SUGGESTIONS FOR CONTRIBUTORS TO SPEECH MONOGRAPHS

A portion of SPEECH MONOGRAPHS is devoted to the publication of articles based on original research, representing the various areas of specialization and techniques of investigation included in the field of speech.

These reports should be given in enough detail to permit the reader to know how the author carried out his investigation and how he arrived at his conclusions. Sources of evidence, conditions of observations, methods of gathering data, should be cited or described. Reports of experimental investigations should follow the usual pattern: statement of purpose, procedures used in gathering data, analysis of data, conclusions. While the inclusion of the data is essential, it should be presented concisely. Tables should be on separate pages. If statistical methods are employed, they should be named, but they need not be described if they are standard, or are adequately explained in some source to which the author can refer.

Articles in SPEECH MONOGRAPHS vary considerably in length. Short ones will be considered by the Editorial Board when they represent investigations the nature or scope of which permits a brief report without omitting essential details. Authors preparing reports of major studies involving extensive data and/or analysis, and anticipating that their material will exceed 8,000 words, should write to the Editor, inquiring about the availability of space and giving an estimate of the probable length of the material.

Contributors should have their manuscripts read by competent critics before submitting them for publication. When manuscripts are based on dissertations, that fact should be indicated, with the name of the director of the original research, in a footnote. In such cases, the director should review the manuscript. All copy should be proofread by at least one person besides the author before being submitted for publication. The number of words in the copy should be noted in the upper right-hand corner of the first page. Titles should be typed in full capitals, with the author's name in capitals, two spaces below. The author's affiliation, school or college (or home city, if he is not so affiliated), should appear, with normal capitalization, two spaces below his name. Footnotes should be typed on separate sheets following the last page of the manuscript. They should be double spaced, with triple spacing between items. Authors should retain a duplicate manuscript, and should send a carbon copy to the editor along with the original. The original copy should be on heavy bond paper.

All manuscripts must conform to the MLA *Style Sheet*, a copy of which can be secured from the Treasurer, Modern Language Association, 100 Washington Square East, New York 3, New York. Price: 25 cents.

Manuscripts will be read by the Editor and members of the Editorial Board before a final decision is made. Accepted manuscripts will be published in approximately the order of their acceptance. The deadline for submitting final revision of accepted manuscripts is two months before date of issue. Publication dates are March, June, August, and November. Contributors should keep the Editor informed of any changes in address.

Reprint orders are sent to authors; orders should be placed at once with the Artcraft Press, Columbia, Missouri. Reprints cannot be secured later.

Manuscripts should be sent to Howard Gilkinson, Editor of SPEECH MONOGRAPHS, 309 Folwell Hall, University of Minnesota, Minneapolis 14, Minnesota.

SPEECH MONOGRAPHS

VOLUME XXIV—No. 1 MARCH, 1957

GREEK AND ENGLISH CONSONANTS

JAMES W. ABEL
Brooklyn College

THIS comparison of modern Greek and English consonants, along with the attendant comments, has grown out of observations made by the author in 1954-55 while teaching English articulation to Greek gymnasium students at Anatolia College, Thessaloniki, Greece.

There are two varieties of modern Greek, [ðimoti'ki],[1] i.e., demotic Greek, the language 'of the people,' and [kaθa'ɾɛvusa], the 'pure' or 'purged' language. The latter is the official language of the country. The former is the language of conversation but it is not restricted to colloquial use; it is also in literary use. Atkinson refers to [kaθa'ɾɛvusa] as ". . . an artificial literary language, based as far as possible upon Attic and Byzantine prose and admitting the vocabulary and constructions of the vernacular in varying degrees."[2] However, as pointed out by Pring, this variety is ". . . not entirely absent from everyday speech."[3] Of particular significance to the subject matter of the present article is the fact that [ðimoti'ki] is the variety of Greek to which a modern Greek grows up; [kaθa'ɾɛvusa] is a variety which he learns in school. Consequently, the educational background of a speaker will affect his familiarity with some linguistic features, including certain phonetic contexts. Most of the examples in this article are drawn from the demotic form, but some are drawn from the learned form, this being the source of the contexts in point.

General observations

Greek has approximately the same number of consonants as English, but each language contains consonants not present in the other. Greek consonants which have only limited, non-distinctive use in English, or are missing from it, are [ɱ], [ɲ], [ʎ], [ç], [x], [ɣ], [ɾ], [ts], and [dz].[4] Greek may be said not to have [ʃ], [ʒ], [h], [tʃ], [dʒ], [w(hw)], and [r]. For all practical purposes, a few Greek consonants are identical with their English counterparts. Most Greek and English consonants are sufficiently similar to be paired and referred to by the same name.

Therefore, except for the use of [ɾ],

[1] Throughout this article Greek words are presented in phonetic transcription rather than in Greek characters or transliteration.

[2] B. F. C. Atkinson, *The Greek Language* (London, 1933), p. 326. See also S. G. Chaconas, *Adamantios Korais* (New York, 1942), Chapter II; Anatol Semenov, *The Greek Language in its Evolution* (London, 1936), pp. 134-6; A. W. Gomme, *Greece* (London, 1945), pp. 24-26 and 115-16.

[3] Julian T. Pring, *A Grammar of Modern Greek on a Phonetic Basis* (London, 1950), p. 6. At this point I wish to acknowledge my general indebtedness to Mr. Pring's writing.

[4] [ts] and [dz] would be removed from this list by those who consider them affricates in English.

rather than [r], to represent the flapped *r* of Greek, for the sake of simplicity, the same phonetic symbols are used to represent both Greek and English consonants in the ensuing discussion,[5] it being understood that the sounds referred to may be somewhat differently articulated in the two languages. These differences are explained in the paragraphs devoted to specific sounds.

As there are similarities and differences in the individual consonants found, so are there similarities and differences in the contextual patterning of the two languages. Both have initial, medial, and final consonants and consonant clusters. Except for those sounds which do not occur in Greek, and vice versa, the initial consonants are the same. All the initial clusters of English are matched in Greek, except for those containing [w] and [ʃ].

Final consonants are less characteristic of Greek than English, many Greek words ending in vowels. Only [-n] and [-s] occur with great frequency. Final [ɾ] is found but much less often than [n] or [s]. Also, final clusters are much less common in Greek. Moreover, the final clusters which do exist are to a greater or lesser degree restricted to learned usage, for example, [xoro'filaks] contrasted with the customary ['filakas] 'policeman.' Elision is a characteristic of spoken Greek.[6] As a result, single consonants are put into positions they do not occupy in isolated words, and more particularly, clusters are created which are not found in word lists. Larger consonant clusters are a distinguishing feature of northern dialects.[7] Nevertheless, it still remains a fact that as compared with English, Greek is characterized by a relative scarcity of final consonants and consonant clusters.

Medially ". . . Greek has an absolute *horror of long consonants.* If two *s* sounds come into contact through combination of words, one of them is simply lost."[8] This may be a slightly dramatic description, but unlike English, Greek does not indicate juxtaposed identical consonants by extra length.

The collective influence of the paucity of final consonants, the restricted use of final clusters, and the avoidance of long consonants can be seen in Greek-English diction. Long consonants may be shortened; members of clusters are often omitted; consonants and consonant clusters are sometimes modified by the addition or insertion of vowels.

Another general observation has to do with assimiliation. In Greek, assimilation creates a pattern of change not found in English. One of the two main features of the pattern is the modification of [n]-plus-plosive combinations in closely linked phrases. Since more than one sound is affected, the changes are summarized on the next page.

The other main feature is the voice assimilation of [s]. Since only one sound is affected here, detailed consideration will be deferred until [s] is discussed.

Finally, Greek makes a less sharp distinction between voiced and voiceless sounds than does English. This gen-

[5] There are five distinctive vowels in Greek, represented in this text by [i], [ɛ], [a], [o], and [u]. For present purposes it is sufficient to say that all are pure vowels as contrasted with English [ɪi], [oʊ], and [ʊu]. Furthermore, in comparison with English, [i] and [u] are more tense, [ɛ] is somewhat lower; Greek [o] is lower than English [o], higher than [ɔ], and farther front than either; [a] is essentially the same in the two languages.

[6] See, for example, Pring, *A Grammar of Modern Greek*, pp. 19 ff.

[7] J. T. Pring, "[gri:k əv mæsidouniə]," *Le Maître Phonétique*, No. 99 (1953), 6. See also Atkinson, *Greek Language*, pp. 314-15.

[8] Henry and Renée Kahane and Ralph L. Ward, *Spoken Greek* (New York, 1945), p. 40.

[n]+[p]>[mb]: [stin 'poli]>[stim 'boli] 'in the city'
[n]+[b]>[mb]: [ðɛn bo'ɾo]>[ðɛm bo'ɾo] 'I can't'
[n]+[t]>[nd]: [ɛn 'taksi]>[ɛn 'daksi] 'O.K.'
[n]+[k]>[ŋg]: [ðɛn ka'talava]>[ðɛŋ ga'talava] 'I don't understand'
[n]+[g]>[ŋg]: [ton 'gɾɛmisɛ]>[toŋ 'gɾɛmisɛ] 'he knocked him down'
[n]+[ɣ]>[ŋɣ]: [ton ɣai'ðaɾon]>[toŋ ɣai'ðaɾon] 'the donkeys' (gen.)

eral native language habit combined with specific factors relating to particular sounds appears to be sufficient reason for the unvoicing which marks the English of Greeks.

How these general differences materialize in the production of specific English sounds and sound combinations will be brought out in the observations which make up the remainder of this paper.

Plosives

[p]: Greek [p] occurs initially and medially in words: [pa'pas] 'priest,' ['plusios] 'rich,' ['pɾaktoɾas] 'agent.' Though found, the initial clusters [spl] and [spɾ] are rare.[9] [p] may be said to occur as a word final. It does so by virtue of the loss of a final vowel: [apo af'to]>[ap af'to] 'from it,' [apo tin ɛ'laða]>[ap tin ɛ'laða] 'from Greece.' In such contexts [p] functions as a medial consonant and is so considered here.[10] [p] may be heard as a word final in interjections and foreign words.[11] The cluster [-ps] occurs: ['kiklops] 'Cyclops.' It may be worth noting that [-ps] is spelled with one letter, [psi] 'psi.'

In Greek, as in English, [p] is a voiceless bilabial plosive, but it is not aspirated. The carry-over of unaspirated [p] into English is most noticeable in the initial position. It is not alone the presence of aspiration which poses a problem for the Greek student of English; there is also the inconstancy of English aspiration.

Since, as has already been noted, in his native language a Greek often assimilates [p] to [b] after [n], he may do likewise in English.

[b]: There are initial and medial, but no final [b]'s in Greek: [bo'ɾo] 'I can,' [lam'bɾos] 'brilliant.' [bl-] and [bɾ-] appear in a limited number of words, some of which, however, are of frequent use: ['bluza] 'blouse,' [bɾos] 'go on.'

Greek [b] is a voiced bilabial plosive, as in English. Unlike English, it is always associated with the Greek equivalent of [n] or [m]. Therefore, English [b] may acquire an [m] on-glide. Greek speakers are also likely to unvoice final English [b], or in voicing it, to add an excrescent [ə].

[t]: [t] is restricted to the initial and medial positions. Initially, it appears alone and in [tɾ-]: [taf'totita] 'identity card,' ['tɾia] 'three.'

Though called by the same name in both languages, distinct contrasts in the fundamentals of articulation separate Greek and English [t]. Greek [t] is unaspirated; English [t] is more or less aspirated depending upon its context. This variability is a major source of difficulty for the Greek attempting English. Having learned that initial [t] before an accented vowel should be strongly aspirated, he may try to say the [t] of [st-] in the same way, and he is almost sure to use too much aspiration

[9] Examples are given in the paragraphs devoted to the first member of the cluster, in this case, those paragraphs about [s].

[10] The same is true of some other consonants.

[11] Subsequently, unless there is particular reason for doing so, interjections and foreign words are excluded from consideration when generalizations about non-occurrence are made.

on absolute final [t] and to think an imploded [t] in this position an error. The very slight aspiration of medial [t] before an unaccented vowel is an even greater hazard. Whereas English [t] is alveolar, Greek [t] is dental, and in the case of some Greeks, the contact is between the blade of the tongue and the back of the upper front teeth, the tip resting at or near the lower gums. In contrast with the open-mouth adjustment and the independent, segmental tongue movement characteristic of English [t] articulation, for Greek [t], during the period of implosion, the mouth is practically closed, and on explosion, simultaneously the jaw drops and the whole tongue, sliding as a unit, breaks contact with the teeth.

In terms of effect upon English articulation, the differences between the manners of articulation are more telling than the difference between the places of articulation. Generally, the sliding whole-tongue release of Greek gives an affricated quality to English [t], which is especially noticeable in [tr-]. More particularly, there is a marked effect upon the flapped articulation of medial [t] before an unaccented vowel, as in *water* and *winter*, due to the absence of the requisite open-mouth adjustment and segmental tongue movement, and because there is no comparable [t] articulation in Greek. These factors, combined with the novel aspiration feature already mentioned, makes this type of English [t] very difficult for Greeks to make. The usual untutored product is [t̪] with far too much aspiration.

Reverting for a moment to the difference between the places of articulation, we find that in their efforts to acquire alveolar placement, some students will over-compensate and make retroflex [t]'s.

As has been remarked generally, [t] following [n] is often assimilated to [d] in Greek. There is a greater or lesser tendency toward the same practice in Hellenic English. Another general observation referred to the intrusion of a vowel into an English cluster. Such a vowel is especially likely to appear between [t] and what should have been a syllabic [n] or [l], as in ['kɑtən] and ['lɪtəl] for ['kɑtn̩] and ['lɪtl̩].

[d]: Greek [d], like [t], is restricted to the initial and medial positions: [du'lapi] 'cupboard,' [a'dio] 'goodbye.' [dɾ-] occurs, but very rarely: ['dɾɛpomɛ] 'I am ashamed.'

Except for the absence of aspiration, there are the same articulatory contrasts between Greek and English [d], as between Greek and English [t]. In fact, an English speaker may have difficulty distinguishing Greek [d̪] from [ð]. As a result, there is good reason for the fact that English [d-], and even more so [dr-], often have unwanted fricative quality when spoken by Greeks. The quickly articulated [-d-] of English does not pose as great a problem as [-t-] because there is no complicating aspiration factor, and also possibly, because of the resemblance between [ɾ] and flapped [d].

Greek [d] is like Greek [b] in that it is always associated with its homorganic nasal. Hence, a Greek may say English [d] with an [n] on-glide.

Final English [d], like final [b], is often unvoiced by Greeks. In the case of [-d] alone, this deviation is frequently made more obvious by an accompanying over-vigorous explosion. Also, sometimes voicing is obtained only at the price of an excrescent [ə].

Intrusive [ə] develops in words like *gladden* and *saddle*, as it does in comparable [t] contexts. The more complicated syllabic cluster [dn̩t] is prone to loss of its [d], making something like

['dɪənt] out of ['dɪdn̩t], for example.

[k]: Greek contains initial [k], not only alone, as in [ka'los] 'good,' but in the clusters [kl-] and [kɾ-], as in [kli'ði] 'key' and [kɾa'si] 'wine,' as well as in [sk-], and rarely in [skɾ-]. There are also medial and final [k]'s: [ka'kos] 'bad,' [ɛk] 'from.' The last is rare, but [ks] is the most common final cluster in Greek: ['pinaks] 'board,' ['naɾθiks] 'narthex.' Like [ps], it is spelled with one letter, [ksi] 'xi.'

Greek [k], as [p], differs from English chiefly in being unaspirated, a fact of which one may be reminded when a Greek essays an initial English [k]. Both English and Greek have palatal varieties of [k], as in *key* and [ɛ'ci] 'there,' respectively. In theory, [c] is used with front vowels only. Actually in Greek, thanks to concomitant vowel changes, [c] sometimes comes before back vowels. Use of [c] before a back vowel in English makes a word such as *coat* sound something like [kjot].

If the Greek's version of [k-] when alone, or more often, in [kr-], is heard as a fricative, there is a reason. Greek has a velar fricative, [x], which like [k] occurs initially both alone and before [ɾ]: [xo'ɾis] 'without,' ['xɾonos] 'year.'

As a preceding Greek [n] many times converts a following [p] or [t] into a voiced plosive, it makes [k] into [g], with the same potential influence on English articulation.

[g]: Initial [g] is rare, occurring in [gl], [gɾ], and alone: ['glava] 'pane,' [gɾem'nos] 'precipice,' ['gaiða] 'bagpipe.' Otherwise, except in [daŋg] 'dong,' it occurs only medially after [ŋ]: [fɛŋ'gaɾi] 'moon.'

This close association with [ŋ] in the native language may show up in the form of a nasal on-glide to English [g]. Thus [g] parallels [b] and [d], as it does in proneness to unvoicing and the development of excrescent [ə] when final. It parallels [k] in having a palatal variant, [ɟ], as in [aŋ'ɟizo] 'I touch,' and a fricative counterpart, [ɣ], as in ['ɣɾiɣoɾa] 'quickly,' which may influence English pronunciations in the same manner as [c] and [x] do.

Nasals

[m]: Initially, it is found alone as in [maθi'tis] 'pupil,' and in [sm-]. Before buses practically displaced street cars, in the two large cities, Athens and Thessaloniki, final [m] was often heard in the foreign word *tram*. Medial [m] comes both before vowels and consonants: [pso'mi] 'bread,' ['pɛmti] 'Thursday.'

Greek [m] is basically bilabial, but medial [m] is subject to modification. It regularly becomes [ɱ] before [f] and [v], as in such as ['ɛɱfasis] 'emphasis' and [no'ɛɱvɾis] 'November,' just as it does for some English speakers in a word like *emphasis*. In Greek, [m] before another consonant may be dropped and the preceding vowel nasalized in compensation: [ðɛm bo'ɾo]>[ðɛ̃ bo'ɾo]. In English, native speakers sometimes do the same, producing such as ['kɑ̃fət] rather than ['kʌmfət]. The difference lies in the fact that in English associated nasality is usually frowned upon whether it comes from native or foreigner. Both in his mother tongue and in English, the Greek may go farther and simply drop [m] before another consonant. Furthermore, final English [m] is very often weakened or dropped.

[n]: [n] occurs in different positions with greater freedom than most Greek consonants and is more like English [n] in this respect than are other Greek consonants like their English counterparts. It occurs alone initially, medially, and finally: [nɛ'ɾo] 'water,' ['mono] 'only,'

[pɾin] 'ago.' [sn-] is found, but rarely. [n] does not join in any final cluster.

It resembles [t] and [d] in being dental and continues the resemblance by being articulated by some with the blade rather than the tip of the tongue. Medial [n] is an unstable sound. As already mentioned generally, [n] is subject to modification by assimilation. Before [p] and [b], it becomes [m], while before [k], [g] and [ɣ] it becomes [ŋ] in phrases which are closely linked. These assimilations may be carried over into English. Medial [n] is also unstable before other consonants, particularly continuants, from where, like [m], it is often dropped. Sometimes the loss is compensated for by nasalization of the preceding vowel; sometimes it is not: [ðɛn 'θɛlo]>[ðɛ̃ 'θɛlo], [ðɛ 'θɛlo]. English [n] is similarly treated before not only continuants but [tʃ] and [dʒ] as well.

[ŋ]: Greek [ŋ] is like English [ŋ] in that it is restricted to the medial and final positions but different in that it appears only in clusters: [aŋgli'ka] 'English,' [ɛŋ'xoɾios] 'native.' The final clusters are, first, [ŋks], which is heard in a few words such as [sfiŋks] 'sphinx,' and which is notable because of its three elements, and second, [ŋg] in the equivalent of *dong*, previously mentioned.

The place of articulation is the same for Greek and English [ŋ], but this does not mean that the Greek speaker always produces [ŋ] as called for in English. In the first place, he finds it very hard to make [ŋ] without an immediately following [g]. Greek [ŋ] is always followed by a velar, usually [g]. Furthermore, Greek [ŋ] is spelled with *gamma*, a letter usually associated with [g] or [ɣ]. Unsophisticated students often appear surprised to find that there is an [ŋ] in Greek. In the second place, [ɲ] may be substituted for [ŋ]. [ɲ] is the Greek variant of [n] before [j]: [panda'loɲja] 'pants.' This substitution is most troublesome in *-ing*. When in such a word as *singing* this substitution is accompanied, as it is likely to be, by certain other practices, the result is a pronunciation which sounds something like *see* plus a prolonged, intermediate nasal with a pitch change at about the midpoint.

The lateral

Greek [l] occurs initially and medially, but not finally: [po'li 'liɣo] 'very little.' There are initial clusters containing [l] to match those in English, but some occur rarely.

From the standpoint of articulation, there are two *l*'s in Greek, a dental, sometimes interdental, [l], which is always clear, or light, and a palatal, [ʎ], which is used before [j]: [paʎj'os] 'old.' Fortunately, [ʎ] does not migrate. The clear [l] does and is usable in English until the final, dark [l] is called for. With this a Greek student has real difficulty. To complicate matters, in his attempts to make dark [l], the student may begin to use a retroflex [l].

Closely related to the production of dark, final [l] is the treatment of the glide between the vowel and [l] in such a word as *school*. Being non-distinctive, the native English speaker ignores it, whereas Greeks are very much aware of it and are inclined to exaggerate the sound. Initial English clusters are also likely to develop an intrusive [ə], as in [bəled] for [bled].

As with medial [t] and [d], so it is with medial [l] before an unaccented vowel. There is no Greek [l] articulation comparable to the very quick, almost flapped [l] of a word like *color* or *yellow*. As a result, [l] in a context of this kind is likely to sound labored and

to become definitely pre-vocalic, as in ['jɛ-lo].

Fricatives

[f], [v], [θ], [ð]: These fricatives are presented together because all are made in Greek essentially as they are in English. All occur in the initial and medial positions: ['filos] 'friend,' [ka'fɛ] 'coffee,' ['vɛvɛos] 'certain,' ['θɛlo] 'I want,' ['laθos] 'mistake,' ['ðoðɛka] 'twelve.' Final [f] occurs rarely, as in certain vocative forms, for example, [vasi'lɛf] 'O, King!', and [-θ] appears in [zɛ'niθ], which though not native is thoroughly naturalized. [f] precedes [l] and [ɾ] in initial clusters, as in [fli'dzani] 'cup' and ['fɾesko] 'fresh,' and [s] in a final cluster, as in [ɣɾa'fɛfs] 'clerk.' [θɾ-] also occurs: ['θɾiaŋvos] 'triumph.'

Hayworth makes special note of the substitution of [d] for [ð] by Greeks in foreign dialect classes at the University of Akron.[12] The writer has not heard this in Greece but he has heard [t] and [d] for [θ] and [ð] from Greeks in the United States.

[s]: Unlike so many other Greek consonants, [s] occurs in all positions: [sas] 'your,' [ɛ'si] 'you.' There are counterparts for all initial English clusters beginning with [s]: ['spoŋgos] 'sponge,' ['splina] 'spleen,' ['spɾoxno] 'to push,' [staθ'mos] 'station,' [stɾa'tjotis] 'soldier,' [sci'li] 'dog,' ['skɾofa] 'sow, slut,' [skli'ɾos] 'hard,' ['smiɾna] 'myrrh,' ['sniçi] 'nape of the neck,' [sfuŋ'gaɾi] 'sponge,' ['slavos] 'Slav.' As has already been stated and exemplified, there are the final clusters [ks], [ŋks], [ps], and [fs].

The contrast between dental and alveolar articulation continues. Consequently, the dental Greek [s] is typically higher pitched than English [s]. Whether this higher pitch is going to sound foreign in the flow of English depends a great deal upon the individual speaking and upon the tolerance of the listener, and to some degree, upon the context. The complex English clusters, such as [-sts], are especially troublesome. That there is a difference between the [s]'s of the two languages is indicated by a bit of rather interesting reverse evidence. In attempting to imitate a native speaker, sometimes Greek students use a fronted [ʃ] for English [s]. Apparently, this is a case of over-compensation for a perceived difference.

The most pervasive problem involving [s] grows out of the assimilation of [s] in Greek, which was referred to in the general observations. [s] before a voiced consonant is often assimilated to a voiced sound: [o 'filos mu]>[o 'filoz mu] 'my friend,' ['pos sas 'lɛnɛ]>['po saz 'lɛnɛ] 'what is your name.' Earlier in this discussion, somewhat detailed consideration has been given to the assimilative changes which often take place in closely linked [n]-plus-plosive combinations. When [ps-], [ts-], or [ks-] follows [n] in such contexts and the plosives become voiced, the [s] becomes [z]: [ton pso'maðon]>[tom bzo'maðon] 'the bakers' (gen.), [stin 'tsɛpi sas]>[stin 'dzɛpi sas] 'in your pocket,' [ðɛn 'ksɛɾo]>[ðɛŋ 'gzɛɾo] 'I don't know.' This tendency of [s] to be easily influenced shows up in English, especially in a very persistent use of [z], or a [z]-like sound, in the initial clusters [sm], [sn], [sl], and [sw] and in medial combinations like those found in *misdeed, housewife, nice boy,* and *this man.* In reverse order, a voiced sibilant is heard in such as *else* and *ounce.*

[z]: Unlike its voiceless partner, Greek [z] does not occur finally, only initially and medially: ['zito] 'long live, hurrah for,' [ma'zi] 'together.' This un-

[12] Donald Hayworth, "The Elimination of Greek Dialect in English," *Quarterly Journal of Speech,* VII (1931), 385-95.

doubtedly contributes toward the strong tendency to unvoice final English [z]. That it is not the entire explanation is indicated by the fact that the medial [z] is also sometimes unvoiced. It will be recalled that one of the general observations made earlier has to do with the less sharp contrast between voiced and voiceless sounds in Greek.

[ʃ], [ʒ]: The defensible generalization may be made that these sounds do not occur in Greek. Sounds approaching them may be heard in such French loan words as *chocolate* and *garage* and as a feature of some local dialects, but they are fronted as compared with the English sounds. As might be expected, the standard substitutions are [s] and [z], respectively.

[tʃ], [dʒ]: As [s] and [z] are used for [ʃ] and [ʒ], the affricates [ts] and [dz] are likely to be substituted for [tʃ] and [dʒ] since they are the nearest Greek approximations to the two English affricates. Both pose special problems when at the ends of words. "Greek [ts] and [dz] occur only at the beginning of a syllable":[13] ['ɛtsi] 'so,' ['dzidzikas] 'cicada.' After what has been said already about the unvoicing of final [d] and [z], the treatment of the [-dz] substituted for [-dʒ] seems fairly obvious.

[h]: The glottal fricative does not occur and is replaced by [ç] or [x], depending upon the following vowel.

Glides

[w]:[14] Greek has no [w]. On attempting this English glide, a Greek speaker may produce anything from an acceptable [w] to a Greek [γ]. The likelihood of this latter substitution increases as the following vowel is higher and farther back. The shibboleth is *woman*. Not only is the tongue raised too high but there is insufficient attention to lip rounding. [w] in clusters is certainly no easier than when alone, and when the preceding sound is another labial, as in *driveway* or *half wit*, there may be some added difficulty with lip rounding.

[r]: The basic *r* sound of Greek is a one-tap trill, or flap. In energetic articulation, it may become a multiple trill. There is neither a vowel *r* comparable to English [ɝ(ɚ)] nor a glide, i.e., [ɹ]. This consonant must be learned as a new sound, and until it is, [ɾ] is usually substituted.

In English, [r] follows [p], [b], [t], [d], [k], [g], [f], [sp], [st], [sk], [ʃ], and [θ] in initial clusters. As has been exemplified, [ɾ] enters into the same combinations except, of course, with [ʃ]. [ʃr-] is a doubly difficult combination for Greeks because both components are new sounds. [θr-], on the other hand, causes little trouble except for occasional trilling of the *r*. After all, [ɾ] or [ř] is used by some native English speakers in this combination. Falling between these extremes, in approximate order of decreasing difficulty, are [(s)pr-], [br-], [fr-], [(s)kr-], and [gr-], all of which are more troublesome than [(s)tr-] or [dr-], because the preceding labial and velar articulations leave the tongue tip freer to tap upward against the roof of the mouth.

[j]: Native Greek [j], as in [pi'jeno] 'I go' and [ji'rizo] 'I return,' often has more fricative quality than is associated with English [j] but it can be converted easily into the English glide.

Summary

Allowing for the error inherent in generalizations, the following summary statements may be made about Greek as compared with English consonants:

1. They are less free to appear finally or to form final clusters.

[13] Pring, *A Grammar of Modern Greek*, p. 15.
[14] No separate consideration of [hw(ʍ)] is considered necessary.

2. Assimilation plays a more consistently telling and accepted role.
3. Voicing is less critical.
4. Voiceless plosives are unaspirated.
5. Dental, rather than alveolar, articulations are predominant.
6. There is no flapped articulation of [t], [d], or [l].
7. Nasals are less clearly differentiated.
8. [ʃ], [ʒ], [tʃ], [dʒ], [h], [w(hw)], and [r] do not occur.

SOME DETERMINANTS OF THE EFFECT OF ORAL COMMUNICATION IN PRODUCING ATTITUDE CHANGE AND LEARNING

DAVID K. BERLO
Michigan State University
and
HALBERT E. GULLEY
University of Illinois

SINCE the introduction of Woolbert's behavioristic approach to rhetoric,[1] most research in speech has operated under a simple stimulus-response schema. Many research problems, however, can not be conceptualized adequately under an S-R rationale, but require a stimulus-organism-response orientation, in which conjunctive relationships between speaker, message, and audience variables are taken into account.

This study, based on data collected in connection with an investigation of intercellular and intracellular speech structure,[2] attempts to isolate some of the variables which operate, either conjunctively or independently, in producing effect. Specifically, this report will consider (a) the congruity hypothesis as a determinant of attitude change; (b) serial position, (c) assertion strength, and (d) predisposition as determinants of learning; and (e) the relationship between learning and attitude change.

I. The Problem

A. *The "principle of congruity" in predicting attitude change.*

Osgood and Tannenbaum recently have reported on a new development in attitude theory, stemming from research in experimental semantics at the Institute of Communications Research at the University of Illinois.[3] Their hypothesis, labelled "the principle of congruity in the prediction of attitude change," can be stated generally:

> Changes in attitude are always in the direction of increased congruity with the existing frame of reference.

Osgood and Tannenbaum suggest that three of the more significant variables affecting attitude change as a result of communication are (a) existing attitude toward the speaker, (b) existing attitude toward the proposition, and (c) whether the speaker supports or opposes the proposition.[4]

Tests of the validity of the congruity hypothesis have employed the Semantic Differential as a measure of attitude.[5] Responses on this instrument can be scored from minus three (maximally unfavorable) to plus three (maximally favorable), with a zero score, represented by a response at the center of the seven-scale interval, defined as a "neutral" position. Under this method of scoring responses, a congruent relationship between one's attitudes toward two con-

[1] C. H. Woolbert, *Fundamentals of Speech* (New York, 1934).

[2] H. E. Gulley and D. K. Berlo, "Effect of Intercellular and Intracellular Speech Structure on Attitude Change and Learning," *Speech Monographs,* XXIII (1956), 288-297.

[3] C. E. Osgood and P. H. Tannenbaum, "The Principle of Congruity in the Prediction of Attitude Change," *Psychological Rev.,* LXII (1955), 42-55.

[4] In their theoretical discussion, Osgood and Tannenbaum do not restrict themselves to consideration of the speaker-proposition relation, but include the condition in which any two concepts are associated in a message.

[5] C. E. Osgood, "The Nature and Measurement of Meaning," *Psychological Bull.,* XLIX (1952), 197-237.

cepts can be defined as the condition in which evaluative responses to the concepts are equally distant from neutral. The nature of the association made between the concepts determines whether, in order to be congruent, the two attitudes are on the same or opposite sides of the neutral point.

If, for example, the speaker favors the proposition, the position of congruency would place attitudes toward both on the same side of neutral; i.e., if Eisenhower praises the Republican party, and a subject's attitude toward Eisenhower is plus three (most favorable), the position of congruency for the subject's attitude toward the Republican party would also be plus three. On the other hand, if the speaker *opposes* the proposition, the position of congruency would place attitude toward the speaker on the opposite side of neutral from the attitude toward the proposition; i.e., if Eisenhower opposes communism, and a subject's attitude toward Eisenhower is plus three, the position of congruency for the subject's attitude toward communism would be *minus* three—equally-distant from neutral, but in the opposite direction.

Before listening to a speech, the attitudes of an individual toward the speaker and his proposition are often not congruent. The "congruity principle" postulates that the hearer, on receiving the message, comes under internal pressure to make his attitudes toward speaker and proposition congruent. The principle of congruity predicts the observable effect of this pressure in terms of both the direction and magnitude of attitude change toward both speaker and concept.

The basis for making a prediction can be best understood by tracing the logical three-step rationale of the principle itself.

1. Let us refer to an individual subject's attitude toward the speaker as "d_s," and his attitude toward the proposition as "d_p" (where "d" represents the distance from neutral, either positive or negative, of the subject's response on the Semantic Differential).

We need to know the "congruent position" of attitude toward both speaker and concept. *The congruent position of the subject's attitude toward the speaker can be defined as that position which would make his attitude toward the speaker congruent with his attitude toward the concept.* As we have seen from the earlier discussion, the algebraic value of this position depends on whether the speaker favors or opposes the proposition. In either case, however, the absolute values (disregarding sign) of the congruent position are equal.

If the speaker favors the proposition, the congruent position of attitude toward the speaker ("d_{cs}") is algebraically equal to the position of the attitude toward the proposition ("d_p"). The same relationship holds for the congruent position of attitude toward the proposition. These relationships can be stated algebraically as:

$$d_{cs} = d_p \qquad d_{cp} = d_s$$

If the speaker *opposes* the proposition, the congruent position of attitude toward the speaker is numerically equal to the attitude toward the proposition, but *opposite* in sign. Again, the same relationship holds for the congruent position of attitude toward the proposition. These relationships can be stated algebraically as:

$$d_{cs} = -d_p \qquad d_{cp} = -d_s$$

2. *The total internal pressure toward congruity exerted on a subject's attitude toward the speaker* ("P_s"), *or on his attitude toward the proposition* ("P_p"), *is equal to the algebraic difference, in attitude scale units, between the atti-*

tude's original position and its congruent position. Symbolically, this relationship can be defined as:

$$P_s = d_{cs} - d_s \qquad P_p = d_{cp} - d_p$$

In other words, the total available pressure exerted on the hearer to change one attitude to make it congruent with the other is determined by the degree of incongruity present between the two attitudes. This pressure is positive when the congruent position of the attitude is more favorable than its prior position, and negative when the congruent position is less favorable.

It can be seen that the total available pressures exertable on each attitude object are numerically equal; however, the actual pressure exerted is determined by the relative polarity of the two prior attitudes. In slightly different terms, we can say that a more favorable or unfavorable (more polarized) attitude will exert more pressure on a less extreme attitude than the less extreme one will exert on the more polarized one. Stated formally:

3. *In terms of producing attitude change, the total pressure toward congruity is distributed between the two attitudes in inverse proportion to their separate degrees of polarization.*

From these three principles, we can derive the specific formulae by which the congruity principle predicts attitude change toward both the speaker and the proposition.[6] If we let "AC_s" and "AC_p" represent the predicted attitude change toward speaker and proposition, respectively, the formulae for prediction of attitude change are:

$$AC_s = \frac{d_p}{d_s + d_p} P_s \qquad AC_p = \frac{d_s}{d_p + d_s} P_p$$

In validating the congruity predictions, Tannenbaum used a source-concept message association which was much more restricted than the speaker-proposition relationship.[7] If it can be demonstrated that this principle is equally valid for the more typical speech situation, it means that one can predict fairly accurately the effects of a speech on the attitudes of the audience, knowing only the prior attitudes toward the speaker and the proposition, and whether the speaker's proposition is in support of or in opposition to the topic. One need know little else about the nature of the message content. To the extent that this principle can be verified, it would seem apparent that it is highly pertinent to rhetorical theory.

B. *Retention of persuasive assertions.*

1. *Serial position and retention.* In any message containing a number of persuasive assertions or factual statements, the major points inevitably are sequentially ordered into one structure or another. The question is often raised as to whether statements in one position or set of positions within a message have a higher probability of being retained than they would if in a different position; i.e., are there optimal serial positions within a message for the transmission of information?

The problem is usually conceptualized in terms of either primacy or recency. With informative messages, the experimental evidence has been inconsistent. Whereas Jersild's study,[8] using eighty biographical statements, supports the primacy hypothesis, Ehrensberger,[9] who

[6] Osgood and Tannenbaum posit a fourth principle, involving a correction for incredulity, described as a positively accelerated growth function. Inasmuch as this function has not yet been explicitly defined, discussion of it is not included in the text of this report.

[7] P. H. Tannenbaum, "Attitudes toward source and concept as factors in attitude change through communications," Unpub. diss. (U. of Illinois, 1953).

[8] A. T. Jersild, "Primacy, Recency, Frequency and Vividness," *Jour. of Experimental Psychology,* XII (1929), 58-70.

[9] Ray Ehrensberger, "An Experimental Study of the Relative Effectiveness of certain forms of Emphasis in Public Speaking," *Speech Monographs,* XII (1945), 94-111.

essentially repeated the Jersild study, and Tannenbaum,[10] who worked with items in a radio newscast, found that those items presented toward the end of the message were better retained.

To the writers' knowledge, no experimental investigation has been reported in which the retention of elements in a persuasive message has been studied as a function of serial position. The present study permits a comparison of the relative efficacy of predictions made under the primacy and recency hypotheses. Specifically, we are interested in an answer to the question:

Can the retention of an assertion within a persuasive message be predicted more accurately by the primacy hypothesis or the recency hypothesis?

2. *Assertion strength and retention.* A similar rhetorical problem concerns the relative retention of assertions, as a function of their presuasive strength. In any persuasive speech, the assertions can be, and usually are, ordered as some function of strength, where the strength of an assertion is defined as the probability that, if the hearer accepts the assertion, he will accept the proposition. The most-used sequential structures are "climax" and "anti-climax."

Two studies, one by Sponberg,[11] and the other by Gilkinson, Paulson, and Sikkink,[12] rank-ordered assertions in both climax and anti-climax sequences. Unfortunately, assertion-strength was confounded with assertion-mass (stronger assertions were allotted more time in presentation); consequently, neither experiment permits an analysis of this problem.

Explicitly stated, the question studied in the present experiment is:

What is the relationship, if any, between the persuasive strength of an assertion used in a persuasive message and its retention, relative to the retention of the other assertions?

C. *Attitudes and learning.*

The area of the inter-relationships of attitudes which individuals have toward a set of concepts and the amount and kinds of information concerning those concepts which they retain from a message, as well as the relationship between the amount and nature of learning and changes in attitude, has not been systematically explored.

If we accept Doob's argument that attitudes themselves are learned,[13] and therefore subject to the same principles of learning as informative material, it would follow that attitudinal and informational responses are both components of semantic meaning. Osgood's notion that attitudes (evaluations) constitute one dimension of semantic meaning supports the kind of argument presented by Doob. Accordingly, it would seem that any one component might have an effect on the nature of the others.

1. *Predisposition and learning.* It has been demonstrated that attitudes affect the way in which stimuli are perceived (e.g., Cooper and Jahoda studies of "Mr. Biggott" cartoons[14]). Lazarsfeld, Berelson, and Gaudet,[15] in studying the mass media habits of individuals of differing political preferences, found

[10] P. H. Tannenbaum, "Effect of Serial Position on Recall of Radio News Stories," *Journalism Quart.*, XXXI (1954), 319-323.

[11] Harold Sponberg, "A Study of the Relative Effectiveness of Climax and Anti-Climax Order in an Argumentative Speech," *Speech Monographs*, XIII (1946), 35-44.

[12] Howard Gilkinson, S. F. Paulson, and D. E. Sikkink, "Effects of Order and Authority in an Argumentative Speech," *Quarterly Journal of Speech*, XL (1954), 183-192.

[13] L. W. Doob, "The Behavior of Attitudes," *Psychological Rev.*, LIV (1947), 135-156.

[14] Eunice Cooper and Marie Jahoda, "The Evasion of Propaganda; How Prejudiced People Respond to Anti-prejudice Propaganda," *Jour. of Psychology*, XXIII (1947), 15-25.

[15] Paul Lazarsfeld, Bernard Berelson and Hazel Gaudet, *The People's Choice.* (New York, 1944).

that, in a situation in which people have a choice of messages, they tend to choose those which are in conformance with their own predispositions. As an exploratory attempt to isolate a possible relationship between learning and attitudinal predisposition, we are interested in the question:

Do individuals of differing prior attitudes toward a proposition differ in the amount of retention of the proposition, the assertions, or the evidence contained in a single persuasive message on that proposition?

Whether or not there are over-all differences in learning attributable to differences in attitude, there may be differences in the learning of specific assertions (e.g., favorable subjects may learn the stronger assertions, while unfavorable subjects retain only the weaker ones). Accordingly, we can ask:

Are there differences among individuals of differing prior attitudes toward the proposition in the retention of specific assertions of differing persuasive strength?

2. *Learning and attitude change.* A second possible dimension of the conjunctive effects of learning and attitudes is the relationship between the amount of information retained after reception of a message and attitude change toward the proposition. Some rhetoricians believe that those individuals who learn the most will be most persuaded. Variations of this hypothesis are at the base of most of the philosophical beliefs concerning the greater power of "truth" in determining policy.

On the other hand, the implications of the "congruity principle" would support the argument that attitude change is not necessarily related to learning.

In light of these opposing kinds of predictions, we are concerned specifically with the question:

Do subjects who change favorably toward the proposition differ from subjects who change unfavorably, or who don't change at all, with respect to over-all retention of the proposition, assertions, or evidence?

II. Results[16]

A. *Congruity hypothesis.*

In testing the validity of predictions made under the principle of congruity, we are interested in two aspects of the data: the direction of change and the magnitude of change.

Given knowledge as to whether subjects were initially favorable, neutral, or hostile toward the speaker and the proposition, and the fact that the speaker supported the proposition, a set of predictions can be generated from the congruity hypothesis as to the direction of change in attitude toward both the speaker and the proposition. The general schema for these predictions is presented in Table I.

One hundred and seventy-four subjects were exposed to the stimulus-message. The congruity hypothesis makes two predictions for each subject, one for change toward the speaker and one for change toward the proposition. Of the 174 predictions made as to change toward the speaker, 117 (67%) were verified. For the proposition, 112 (64%) were verified. In both cases, the number of correct predictions departs significantly from chance (P .0001). The congruity hypothesis successfully predicted the direction of attitudinal effects as a consequence of exposure to this message for about 2 out of every 3 subjects.

Predictions were also made, from pretest scores, as to the exact scores which would be obtained after the message had been received (see discussion of the Problem). Pearson product-moment correlations were computed between predicted

[16] The procedures used in collecting the data of this study have been presented in detail in a previous report. See Gulley and Berlo, *op. cit.*

TABLE I
SCHEMA FOR DIRECTIONAL PREDICTIONS MADE BY THE PRINCIPLE OF CONGRUITY

Predisposition toward the speaker	Predisposition toward the Proposition: Favorable		Neutral		Hostile	
	Source Ch	Concept Ch	Source Ch	Concept Ch	Source Ch	Concept Ch
Favorable	+	+	o	+	—	+
Neutral	+	o	o	o	—	o
Hostile	+	—	o	—	—	—

and observed post-test scores. The correlation obtained for attitude toward the speaker was .63; for attitude toward the proposition, .71. Both of these correlations are highly significant (P .001), indicating that the congruity hypothesis enabled predictions as to magnitude of change with well above chance success.

B. *Retention of persuasive assertions.*

1. *Serial position and retention.* This section of the analysis is concerned with the relative efficacy of the primacy and recency hypotheses in predicting the learning of assertions as a function of their serial position.

The five assertions used in the message were presented under three different sequential structures: climax, anti-climax, and pyramidal. Under these conditions, the same assertion was sometimes placed in different serial positions. The serial sequencing under each of the three experimental treatments is shown in Table II.

If we are to control for possible differences among assertions of differing persuasive strength, comparisons can be made only for one assertion at a time. Under this procedure, twelve comparisons can be made as to the relative superiority of primacy and recency. The actual predictions which were made, as well as a comparison of the predictions with the obtained data, are presented in Table III.

TABLE II
ORDER OF PRESENTATION OF EACH OF FIVE ASSERTIONS UNDER THREE EXPERIMENTAL CONDITIONS

Assertion Strength	Serial Position: Climax	Anti-climax	Pyramidal
I	5	1	3
II	4	2	4
III	3	3	2
IV	2	4	5
V	1	5	1

Ten of the twelve predictions made under the recency hypothesis were verified. This result is significant beyond the 5% level of confidence. In this experiment at least, the recency hypothesis furnishes the better basis for prediction of the relationship betwen serial position and assertion retention.

2. *Assertion strength and retention.* In performing this analysis, the writers originally intended to control for serial position; however, the results were so strikingly clear that this precaution was not thought to be necessary.

The percentage figures presented in Table IV represent the number of subjects in each experimental group who correctly recalled each of the assertions, as well as a summary for all experimental subjects.

Chi-square analyses of the frequency data, performed separately for each experimental treatment, and for the totals, were all highly significant. There are large differences in retention which are related to differences in the persuasive strength of the assertion.

Inspection of the results reveals that these differences follow a consistent pat-

TABLE III
PREDICTED AND OBSERVED EFFICACY OF THE PRIMACY AND RECENCY HYPOTHESES IN EXPLAINING THE RELATIONSHIP BETWEEN SERIAL POSITION AND RETENTION OF EACH OF FIVE ASSERTIONS

Assertion Strength	Predicted Relationships*		Observed Results	Number of Verifications	
	Primacy Hypothesis	Recency Hypothesis		Prim.	Rec.
I	A>P: A>C: P>C.	C>P: C>A: P>A.	P>C>A.	1	2
II	A>P: A>C.	C>A: P>A.	P>C>A.	0	2
III	P>A: P>C.	C>P: A>P.	C>P>A.	1	1
IV	C>A: C>P: A>P.	P>A: P>C: A>C.	P>A>C.	0	3
V	C>A: P>A.	A>C: A>P.	A>C>P.	0	2
				2	10**

*"A" represents anti-climax; "C," climax; "P," pyramidal.
**Significant at .038 level, two-tailed test.

TABLE IV
PERCENTAGE FIGURES FOR SUBJECTS WHO RECALLED EACH OF THE FIVE ASSERTIONS IN EACH OF THE THREE EXPERIMENTAL CONDITIONS

Assertion Strength	Experimental Conditions			
	Climax	Anti-climax	Pyramidal	Total
I	28%	18%	31%	26%
II	30	26	31	29
III	52	31	43	42
IV	38	40	57	45
V	81	84	80	82

tern. With only one reversal (assertions three and four in climax structure) and one tie (assertions one and two in pyramidal structure), retention of the assertions is a negative monotonic function of strength.

It can be observed from the frequency data for the entire experimental group that if we rank order assertions as to frequency of recall, the obtained rank order is the inverse of the one obtained from ranking the assertions as a function of strength.

C. *Attitude and learning.*

1. ***Predisposition and learning.*** As this problem was originally stated, we were interested in the relationship between a subject's attitude toward the proposition prior to reception of the message and the amount of material (proposition, assertions, and evidence) the subject retains from a persuasive message on that proposition. In order to investigate this problem systematically, it would be desirable to have a proposition toward which prior attitudes are distributed all along the continuum. Regrettably, though a pilot study indicated that attitudes toward this proposition were rectangularly distributed, the responses secured in the experiment are restricted as to range of prior attitude. Only 14 subjects responded unfavorably to "general education courses" on the pretest. For this reason, the analyses have been restricted to a comparison of "favorable" and "non-favorable" groups, in which "non-favorable" in-

cludes both neutral subjects and the fourteen unfavorable subjects.

Responses of subjects to the question on recall of the proposition were categorized into two groups: "did recall" and "did not recall." A chi-square, computed on the number of favorable and non-favorable subjects who did or did not recall the proposition, was insignificant ($\chi^2 = 1.46$, df = 1).

In an attempt to refine the sensitivity of the analysis, comparison was made among subjects of three degrees of favorableness toward the proposition, eliminating the non-favorables. The degree of favorableness toward the proposition did not differentiate between subjects who were or were not able to recall the proposition ($\chi^2 = .60$, df = 2).

Two similar analyses were performed on scores obtained from the item measuring recall of the five assertions. The chi-square computed on frequency data for favorable and non-favorable subjects was not statistically significant ($\chi^2 = 4.21$, df = 3). The chi-square resulting from a comparison of subjects of differing degrees of favorableness was also not significant ($\chi^2 = 3.5$, df = 6).

A chi-square, computed on frequency data for learning of the evidence by favorable and non-favorable subjects was insignificant ($\chi^2 = .60$, df = 2), as was the chi-square computed on the three groups differing in degrees of favorableness ($\chi^2 = 3.94$, df = 4).

In addition to a comparison of prior attitude toward the proposition and over-all learning of the proposition, assertions, and evidence, we were also interested in possible differences between favorable and non-favorable subjects with respect to learning specific assertions. In order to maximize any possible differences, the comparison presented in Table V was restricted to the 35 most favorable subjects and the 39 least favorable. Inspection of the table reveals that these two groups did not differ with respect to learning of any of the five assertions.

TABLE V

COMPARISON OF THE NUMBER OF SUBJECTS IN THE FAVORABLE (N = 35) AND UNFAVORABLE (N = 39) GROUPS WHO WERE ABLE TO RECALL EACH OF THE FIVE ASSERTIONS

Assertion Strength	Predisposition toward the Proposition: Fav. Group f	Fav. Group %	Unfav. Group f	Unfav. Group %
I	11	31%	13	33%
II	10	29	13	33
III	16	46	15	38
IV	13	37	18	46
V	30	86	28	72

These analyses definitely do not support a hypothesis that the amount of information retained after exposure to a message is dependent on the initial attitudes toward the proposition.

2. *Learning and attitude change.* If attitude change is related to learning of the material presented in the message, we would expect learning scores for subjects who changed favorably toward the proposition to differ from scores of those who changed unfavorably, or those who did not change at all. A chi-square, computed on frequency data presented in Table VI, was significant at the .02 level. Of those subjects who were not able to recall the proposition, 75% changed favorably toward the proposition, while 25% did not. Subjects who changed favorably were less likely to recall the proposition than were subjects who did not change favorably. No

TABLE VI

CHI-SQUARE ANALYSIS OF FREQUENCY DATA FOR CHANGE TOWARD THE PROPOSITION AND RECALL OF THE PROPOSITION

Direction of Change	Recall of the Proposition: No	Yes	Total
Favorable	82	28	110
Unfavorable	28	22	50
			160

$\chi^2 = 5.51$,* df = 1

*Significant beyond .02 level.

logical explanation for this significance offers itself. In light of the fact that this is the only significant chi-square of several which were computed, it is quite possible that the significance can itself be ascribed to chance.

Subjects who changed favorably toward the proposition were compared with those who did not, with respect to recall of the assertions. The chi-square which was computed was not significant ($\chi^2 = 5.64$, $df = 3$). A similar analysis compared subjects who changed favorably and those who changed unfavorably toward the proposition, with respect to recognition of the evidence. The chi-square was not significant ($\chi^2 = 3.07$, $df = 3$).

The results of this study would indicate that the direction of change of attitude toward the proposition is not related to learning of the assertions or the evidence.

III. Discussion

The "congruity principle" of Osgood and Tannenbaum was shown to be a highly significant predictor of attitude change toward both the speaker and the proposition. The fact that this principle can be verified in what approximated the public speaking situation has definite implications, both methodological and philosophical, for rhetorical theory.

To the extent that this is an explanatory principle of speech effect, it indicates that the success of the speaker in influencing an audience to change their attitudes toward either a proposition or the speaker himself rests to a large extent on what the audience's prior attitudes are toward him and toward his proposition. It reinforces the traditional importance of "ethos" in the speaker, inasmuch as the favorableness with which he is viewed is of prime importance in securing acceptance of a substantive point of view.

Secondly, the "congruity principle" is an empirically-based denial of the so-called "rational" principles of rhetoric. It suggests that an audience expects the speakers whom it highly favors to praise only ideas which it also favors, and to attack only ideas which it opposes. Failure on the part of the speaker to adhere to this expectation will lower his own stature with the audience. What this implies in the practical speech situation is that whenever a speaker feels required to defend an unpopular belief, he can assume that he will suffer for it in popular appeal. Also, to the demagogue who cares not what he supports, but only that he himself gains support, it clearly indicates the road to popular success. In fact, the congruity principle supports the "followership" theory of leadership; i.e., the successful leader is the individual who advocates only what has already been accepted by the public, and opposes only what has been rejected.

Another implication of this principle involves a further extension of the notion that acceptance is achieved through association with popular ideas. If the speaker is to increase his own public prestige, he must not only praise what the public approves of, but must restrict his praise to those ideas which are approved *to a greater extent than he is approved,* and criticize only those ideas which are opposed more than he is approved. In other words, not only will a speaker lose support by opposing popular ideas, but he will be perceived less favorably by praising ideas which, though popular, are not as popular as he is, and also by criticizing ideas which, though unpopular, are not as unpopular as he is popular.

Such a hypothesis is in line with the

political credo to "praise motherhood" and "damn sin;" however, the range of situations to which the congruity principle applies would seem much wider than this simple witticism implies.

The extent of generalizability of the congruity hypothesis remains to be determined. If, as present data would tend to indicate, it can be effectively used in many speech situations, the serious philosophical and methodological implications are obvious. No serious student of the use of language in social control can safely ignore these possible ramifications.

To the extent that attitude change is determined by the effect of the speaker-proposition association and prior attitudes toward each, we would not expect message content, organization, style, delivery—in short, the major variables in the teaching of public speaking—to differentiate between successful and non-successful speech. In the context of the present study, the observed lack of any meaningful relationship between prior attitude and learning, and between learning and attitude change, is consistent with the observed significant effect of "congruity." If effect is greatly determined by the latter principle, it would be expected that the amount of factual information retained would not be particularly important. In other words, the lack of statistical significance found in comparing learning and attitude change is in itself indirect supportive evidence for the importance of the congruity principle.

Additional information on the generalizability of this principle should prove of great interest. It is possible that it will not hold in all, or even most, speech situations. If the latter is the case, it should prove beneficial to explore systematically the conditions under which it will and will not predict change. And, of course, under the most favorable conditions, it does not by any means account for all of the observed effect. Undoubtedly, many of the traditional communication variables are operating. Nevertheless, this kind of research has great pertinence to the type of rhetorical theory which can be constructed, to the kinds of techniques which are used in teaching speech skills, and probably most importantly, to the kinds of problems which it poses for a philosophy of rhetoric, and the emphases which need to be placed on the teaching of listening.

Let us turn our attention to another aspect of the present study. The evidence from this experiment tends to support the recency hypothesis as an explanation of information retention from an orally delivered message. The present data are in agreement with the research of Ehrensberger and Tannenbaum, although in conflict with the Jersild study. The contradiction of experimental results prohibits the statement of a generally applicable principle. The writers feel that attempts to demonstrate a universal superiority of one or another learning hypothesis in communication theory will probably not be successful.

Of final interest is the observed relationship between the strength of an assertion and the level of its retention. Retention was found to be negatively related to strength. It hardly seems likely that such a finding reflects a causal relationship between these two variables, though of course this is possible. Rather, it would seem that the observed relationship could be best explained by finding a third variable, related to both strength and retention, which can explain the correlation.

In an attempt to isolate such a variable, the writers collected some additional data from another sample of the same population as the experimental sample. Fifty subjects were given a list

of 25 assertions on general education courses, including the five which were used in the experiment. Subjects indicated how frequently they had heard each of the assertions, by checking one of four categories: frequently, occasionally, seldom, or never.

It was found that strength and familiarity were positively related; in fact, the five assertions were placed in the same rank order on both variables. In light of this, the observed strength—retention relationship can not be explained by familiarity.

It might prove interesting to test the hypothesis that learning of the assertions is positively related to the *novelty* of the assertions, which might be negatively related to assertion strength. In other words, we might predict that subjects will tend to learn best that which is novel, and will tend to evaluate those assertions with which they are most familiar as "strong." In fact, perception of strength might be causally determined by familiarity. The tenability of this hypothesis will rest on the results of further study.

A STUDY OF THE CHARACTERISTICS OF AN AUDIENCE VIEWING A SPECIFIC TELECOURSE

JOHN R. SHEPHERD
University of Washington

THE instructor who teaches via television is faced with the practical problem of analyzing his audience. It is important, for example, for him to have some idea of the scholastic background of his pupils. There are a number of demographic factors which could be important in terms of the general approach he might take in his teaching. Finally, there could be behavioral characteristics of the group, in terms of their use of, and attitudes toward television which would have some bearing both on the material and the method of teaching. While the study reported here does not pretend to provide conclusive evidence on any of these problems, it can be considered as a preliminary step in that direction. With additional evidence that will be available as the result of other studies now in progress, which developed from this one, we may be in a position eventually to give the television instructor a definitive description of his group. Meanwhile, the results reported here should be helpful in throwing light on some of these problems, but should obviously not be considered as more than that.

With the above limitations in mind, the major problem of this study can be stated in the following form: What are the characteristics of that portion of the audience viewing the "Intermediate Algebra" telecourse, as broadcast by Station KCTS-TV, which identified itself either through the purchase of the *Viewer's Guide* and/or enrollment in the corollary Home Study course?

Previous Studies. Most of the studies reported in the literature which deal with the question of the audience and educational television mention specific characteristics of that audience only incidentally. An exception to this, however, was the study reported by James E. Lynch in a recent issue of the *Speech Monographs*.[1] This researcher analyzed the set owners in the Detroit area who reported having viewed the program entitled "University of Michigan Television Hour." The sample was drawn from 764 households used in the 1954 Detroit Area Study. According to this survey, 87% of the homes in that area were equipped with television, and of these, 50% reported having seen the particular program being studied. Twenty-four percent of that number said they viewed the program regularly, which was defined as "one to four times a month." Lynch analyzed his data into five categories, comparing the viewers with the non-viewers. Age, education, sex, income, and occupation were the bases of his analysis. Of importance to the present study are the first four of these. It was found that all age groups up to sixty viewed the program. Those who viewed were more likely to have had some college training than those that did not. The amount of viewing, at least among the "regular" viewers, was slightly higher for men. Finally, it was observed by Lynch that in homes where the principal wage-earner earns over $5000, there was

[1] James E. Lynch, "An Educational Television Audience," *Speech Monographs*, XXIII (1956), 60-65.

a "slight tendency" to view the program more.

More typical of the usual approach to the problem, however, was a study done at Western Reserve University.[2] Here, by means of 661 interviews, an attempt was made to assess the size of the audience using various telecourses presented by that University. In this process it was found that most of the respondents were women.

Another study, which dealt with the question somewhat more directly, was done by Evans, Roney and McAdams.[3] These researchers, making use of the telephone interview method, contacted 384 subjects "randomly selected" from the Houston area and asked a series of questions related to the extent of use, size of audience, program preferences, and suggestions for future programs.

James D. Finn, in his article entitled, "Television and Education: A Review of Research,"[4] confined himself, either by choice or necessity, to reviewing research concerned with the questions of the effectiveness of television as related to teaching, and again only incidentally does he touch upon research pertinent to this study.

However, if there has been little research on the specific characteristics of those audiences viewing educational television, there has been no such lack in the study of what we might call the "general" television audience. Such research is of importance here for purposes of comparison between the selected audience and the general one.

The most recent, and perhaps most comprehensive report of such research is to be found in T. E. Coffin's "Television's Impact on Society."[5] This review summarizes in considerable detail both past and current research on various types of audiences using television, and although it is pointed out that there have been studies of such specialized audiences as "daytime TV viewers," "summertime viewers" and the like, no mention of specific studies of the educational television audience is found. This can be accounted for, perhaps, in Coffin's remark that "Such specialized audiences tend to display the fundamental characteristics of T.V. owners at large, modified here and there by factors associated with specific viewing behavior."[6]

Finally, two authors, Wischner and Scheier writing in the *American Psychologist*[7] point up the need for research into the composition of such particular audiences as the one under examination here when they say:

> "We . . . need information on the types of students for which television is most effective in terms of aptitude and background characteristics.
>
> . . . Very little is known about the background factors that may make an individual more or less receptive to T.V. instruction."[8]

Although this study does not attempt to answer directly the questions raised by these authors regarding the receptiveness of the audience and effectiveness of television, their remarks do suggest the need for basic research into the background characteristics of educational tel-

[2] William K. Cumming, "This is Educational Television," 1954. (Lithoprinted by Edwards Brother, Inc., Ann Arbor, Michigan), 127.

[3] Richard I. Evans, H. Burr Roney and Walter J. McAdams, "An Evaluation of the Effectiveness of Instruction and Audience Reaction to Programing on an Educational Television Station," *Journal of Applied Psychology*, 39 (1955), 277-279.

[4] James D. Finn, "Television and Education: A Review of Research," *Audio-Visual Communication Review*, 1 (1953), 106-126.

[5] Thomas E. Coffin, "Television's Impact on Society," *American Psychologist*, 10 (1955), 630-641.

[6] *Ibid.*, p. 362.

[7] George J. Wischner and Ivan H. Scheier, "Some Thoughts on Television as an Educational Tool," *American Psychologist*, 10 (1955), 611-614.

[8] *Ibid.*, p. 614.

evision audiences as the beginning of the approach to the larger and perhaps more important problem.

I. Sample and Procedure

The Telecourse. The telecourse, "Intermediate Algebra," was based on a similar course offered as Mathematics 101 in both the regular day and evening classes of the University of Washington. The most complete description of the course and its content is to be found in the University Evening Class catalog, and it reads as follows:

> This course presupposes one year of high school algebra. It covers the material included in high school Algebra III. Factoring, fractions, systems of linear equations, theory of exponents, and radicals, quadratic equations, ratio, proportion and variation, the binomial theorem, logarithms. Prerequisite, one year of high school algebra. Not open for credit to students presenting one and one-half years of high school algebra.[9]

All programs were telecast over the educational station in Seattle, Washington, which is KCTS-TV, Channel 9, VHF. All programs were transmitted at 7:30 p.m., Monday, Tuesday, and Wednesday of each week for a period of twenty weeks, beginning January 3, 1956. They consisted of a "correction session" on Monday, featuring the review and analysis of work submitted during the previous week, and lectures delivered by the head of the Mathematics Department on the two remaining programs.

Three types of "participation" by students were possible. First, a student could "officially" register for the course. For this, the viewer paid the regular Home Study fee of $30.00, was required to buy the text, complete the weekly assignments, and to take the examinations. The second type of participation was that called "regular" participation. The viewer was placed into this category if he purchased the *Viewer's Guide,* a work book prepared by the lecturer and sold by the University for $1.00. The "regular" participant was not obliged to buy the text, nor was he required to send in the regular exercises and examinations. Finally, of course, it was assumed that there were viewers who did not fall into either of these groups, but who did constitute another kind of participant.

Selection of the Sample. The sample for this survey was drawn from the first of the two groups named above: those who were either "officially" or "regularly" enrolled in the telecourse. It was assumed that the individual interested enough in sending for the *Guide* and/or enrolling in the course would likely be a relatively constant viewer, and hence his response would be pertinent to the study at hand. In addition, this procedure made it possible to delimit the number to whom schedules should be sent. No effort was made to sample that portion of the audience not included in these groups. Twenty-two students made up the first of them, while 291 were included in the second.

The schedules were mailed during the third week of the course. Two weeks later, the follow-up questionnaire was sent to those not responding to the original one.

Although no effort was made to include the third type of participant in the sample, an effort was made to assess the number of people who might be in that group, by means of offering to anyone who requested it a free mid-term examination. These tests were offered in the sixth week of the course and the offer was announced on three separate broadcasts. It was assumed that only rather regular and interested students would respond to such a request. Of the 80 received in response to the offer, 38 of them were from individuals already

[9] University of Washington *Evening Classes Bulletin,* Spring 1956, pp. 31.

included in the sample. The remaining 42 were not previously accounted for. The discrepancy between the requests and previous inclusion in the sample should suggest to the reader the limitations of the present study. How representative the sample of the present study may be is a matter for conjecture.

The Schedule-Construction and Analysis. In consultation with the Departments of Mathematics and of Correspondence Study, two forms of the schedule were prepared. Form I was designed for the "regularly" enrolled students (credit), and Form II was devised for those who purchased the *Viewer's Guide* only (non-credit). The principal difference between them was that on Form I three questions were asked regarding past and current enrollment in the University, and on II, two questions were included that related to previous telecourse viewing. Form I contained a total of twenty-one questions. Form II had eighteen.

The two sets of data were first tabulated separately, but upon inspection it was found that little difference was observable between the two groups. Therefore, wherever it was possible to do so, the results were combined and presented in that way in this report. Unless otherwise noted, all tables and analyses represent the combined results.

II. Results

Response to the Questionnaire. From the total of 313 questionnaires mailed, a response of 81.8%, or 256, was obtained. Two things may account for this relatively high rate of return. First, it may be that this sample represents a special and particular group who were especially motivated to return the questionnaire. Second, it may be that the viewers responded because they were reminded by repeated announcements on the telecourse to return the questionnaire.

Sex. Two hundred fifty of the respondents indicated their sex. Of this number, 170 were men, 80 were women. This finding is in disagreement with those of other researchers. A study conducted at Western Reserve University found that the audience is "mainly women" for educational television programs.[10] The University of Omaha, reports that research there found such audiences to be "predominantly women". . .[11] In both of these cases the audiences sampled were for telecasts of programs whose subjects were not science or mathematics. Lynch reports that in his sample men tended to view the program "slightly" more than did women but did not define the subject of the program(s) they viewed.[12] The implication is that possibly generalizations regarding sex of audiences should be qualified in terms of the particular subject being presented.

Age. The audience was asked to indicate their birth dates, from which the data on ages were obtained. Table I presents these findings, with a comparison of the age of a sample of the students enrolled in the Extension (evening) Classes for 1956.[13]

It is interesting to note the differences between the figures for the second, third and fourth intervals in the age groups compared. The differences here indicate that perhaps the telecourse tends to attract an older segment of the population than does the traditional evening class.

Elroy L. Stromberg, writing in the *American Psychologist,* observed that the age range he found for students enrolled

[10] Cumming, *op. cit.,* p. 134.
[11] Cumming, *op. cit.,* p. 127.
[12] Lynch, *op. cit.,* p. 60.
[13] These data were collected in a survey of Extension Classes conducted by the Extension Class Department of the University of Washington in 1956.

TABLE I
AGE OF VIEWERS AS COMPARED TO STUDENTS ENROLLED IN UNIVERSITY EXTENSION CLASSES

Age Group	Telecourse Viewers Number	Telecourse Viewers Percent	Extension Class Students Percent*
0-19	29	11.8%	5.1%
20-25	19	7.8	30.5
26-35	57	23.3	34.3
36-45	79	32.3	16.9
46-55	34	13.9	9.3
56-up	27	11.0	3.9
	N=245	100.0%	100.0%

*These percentages are based on a sample of 2595 drawn from the 3696 students enrolled either for credit or non-credit in the University of Washington Evening Class program.

in a telecourse was "19 to 50."[14] Unfortunately, Stromberg does not give a full distribution of this finding so direct comparison is not possible. Lynch, on the other hand, reports finding that viewing is about the same for all age groups, excepting for those over sixty where a drop in viewing occurs.[15] In regard to the general television audience, Coffin states that the research indicates television owners "tend to cluster in the intermediate age groups" and that "53.3% of the television housewives are between 25 and 44."[16] Whether or not "housewives" and "television owners" are comparable to the sample of the present study is a question. Assuming that they are, then it would seem that this educational television program is indeed reaching an older audience. If true, the implications for education are great.

Academic Background for Course. Because of the course content of Intermediate Algebra, it was desirable to know whether or not the respondent had had a similar course in algebra. Because of the stated requirements, and because of the review nature of the first weeks of the course, it is not surprising to find that of the 228 responses to this item, 128 or 56.1% of them were affirmative.

Source of Information About the Course. Where the viewers heard about the course is important to any administrator planning the advertisements for programs. In addition to the Home Study Bulletin, which carried information on the course as a Home Study Course rather than as a telecourse, the Department of Correspondence Study issued special printed announcements. These, along with five other alternatives, were offered to the respondent as the source of his information. Table II reports these findings. Because of multiple responses, the number of them exceeds the size of the total sample.

TABLE II
SOURCE OF INFORMATION ON INTERMEDIATE ALGEBRA

Source	Number	Percent
Teacher	9	2.9%
Friend	20	6.4
Newspaper	109	34.6
KCTS	87	27.6
Home Study Bulletin	23	7.3
Print. Annct.	50	15.9
Other	17	5.4
	N=315	100.0%

Education. In order to ascertain the educational level of respondents, they were asked to indicate the highest grade completed in school. The results from this inquiry are reported in Table III.

[14] Elroy L. Stromberg, "College Credit for Television Home Study," *American Psychologist*, 7 (1952), 507-509.
[15] Lynch, *op. cit.*, p. 56.
[16] Coffin, *op. cit.*, p. 631.

TABLE III
HIGHEST GRADE COMPLETED

Grade Level		Number	Percent
Elementary:			
	4	1	.4%
	5	—	—
	6	2	.8
	7	2	.8
	8	10	4.0
Secondary:			
	9	12	4.8
	10	14	5.6
	11	18	7.2
	12	103	40.4
College:			
	Fresh.	16	6.4
	Soph.	15	6.0
	Junior	6	2.4
	Senior	35	14.1
Graduates:			
	M.A.	3	1.2
	Ed.D	—	—
	Ph.D	1	.4
	Other	11	4.4
		N=249	100.0%

On the basis of the data in Table III, it is suggested that the common hypothesis, that those who view educational television programs are better educated than ordinary viewers, may be true.

When the figures of Table III are compared with those reported by Coffin for the educational level of "household heads" in television homes, it is found that differences do exist. He notes that 24.0% of the general audience have less than a grade school education; 24.9% have "some high school"; 34.4% have completed high school, and 16.7% have "some college or more."[17] Coffin does not include, however, the number of cases used in arriving at these percentages. Lynch, on the other hand, concludes on the basis of his analysis, that ". . . college-educated persons are more likely to be viewers. . . ."[18]

Employment. Of the 244 responses to this item, 69.7% of them indicated the individual was employed. If the respondent were a student, he was asked to indicate the grade in which he was enrolled. This question produced 32 responses, which constitute 12.5% of the *total* sample. Five of these students were in the tenth grade, seven in the eleventh grade, nine in the twelfth grade, and six were freshmen in college. There was some indication that there should have been a provision for the respondent to indicate a "retired" status, and the failure to provide for this in the schedule may account for some of those who indicated they were "unemployed."

Purpose of Viewing. To assess the motive(s) for viewing a telecourse is beyond the scope of this present study. However, one question was included in the schedule which may at least suggest one of these. The response to it may also indicate an attitude the specific group held toward the particular telecourse. The viewer was asked to indicate whether or not he felt the telecourse would help him in his "professional advancement." Obviously loaded in the direction of a positive answer, it is not surprising that 67.1% of the total response of 225 replies were positive. The bias of the question makes extended analysis dangerous. It is enough to say that the response seems to be consistent with the remainder of the data in that it suggests a specific and explicit purpose in viewing the course.

Approximate Annual Income. The respondents in this study were asked to indicate their nearest approximate annual income on an eight point scale, ranging from $2000 annually to $9000 or more. The results of this inquiry are reported in Table IV.

Referring again to Coffin's report, it is suggested that the income of the present sample is higher than that of the general television audience. Coffin's data show that 39.4% of the TV homes in the general audience have an income of

[17] Coffin, *op. cit.*, p. 632.
[18] Lynch, *op. cit.*, p. 60.

TABLE IV
NEAREST APPROXIMATE ANNUAL INCOME

Amount	Number	Percent
$2000	26	12.5%
3000	23	11.1
4000	32	15.4
5000	62	29.8
6000	23	11.1
7000	24	11.5
8000	8	3.9
9000	10	4.8
	N=208	100.0%

$5000 or more.[19] Comparing this figure with those data taken from Table IV, it is found that 61.1% of the viewers in the present study report an annual income of $5000 or over. Referring again to the work of Lynch, he found that there was a "slight tendency" for those people living in homes with incomes over $5000 to view the Michigan program more.[20]

Competition for the Television Set. Competition within the family for the use of the set might be one factor which would limit the use made of it for viewing educational programs. If this were a factor in hindering the effective use of the program, one would expect to find the majority of the telecourse viewers with only single sets, and heavy competition from other members of the family for the use of it. On the basis of the data obtained in this survey, both of these conditions do prevail. It was found that 227 of the 249 respondents to the question stated that they had only one set. When asked the additional question of how many others in the home used the set, it was found that 79.2% of the 236 answering reported they shared the set with from one to three persons. Eleven and nine-tenths per cent share it with four, while the remaining 8.9% must fit their viewing to accommodate from five to six or more individuals.

[19] Coffin, *op. cit.*, p. 631.
[20] Lynch, *op. cit.*, p. 60.

Group Viewing. One possibility suggested by the findings reported above is that a large amount of group viewing was in progress. In order to have some check on this, an analysis was made of the requests made for the mid-term examination. It was found that of the 80 requests, 50 were for single copies, while the remaining 30 asked for multiple copies. This latter group requested a total of 85 copies. The range of these requests was from two to fifteen; the mean number requested was three. These data seem to suggest that group listening did occur, but how widespread this practice was is difficult to ascertain with the available information.

Approximate Time Spent Viewing Television. A seven interval check item was included in the schedule in order to obtain information on the approximate number of hours that were spent in viewing television daily. It was found that of the 234 answers to this question, 58.1% of these indicated the respondent viewed from two to three hours per day. In connection with this, it should be pointed out that the category "one hour" was inadvertently omitted from the choices provided the respondent, and hence the results may be skewed. However, this finding does compare with that found for the general television audiences, and may not be as distorted as one would expect. The "Videotown" study[21] shows that time spent viewing varies from 2.41 hours to 3.83, depending on the member of the family consulted. Coffin says that the literature "commonly" reports "upwards of 2 or 3 hours per day."[22]

General Program Preference. In order to estimate the preferences of this sample, the respondent was asked to rank

[21] Herbert L. Marx, Jr., "Television and Radio in American Life," Vol. 25, No. 2, 1953, p. 198.
[22] Coffin, *op. cit.*, p. 632.

six common program types. By tabulating the number of ranks given each category and averaging each of these, it was possible to assign a rank order to the six. Using this procedure, it was found that the program types were ranked by 203 respondents as follows: drama, sports, variety, news, education, and music. No effort was made to define the types—the list was provided without explanation.

At least one study has investigated the preference of telecourse viewers for certain types of programs. Evans,[23] asked his sample to state preferences for seven types of programs. Of these 83% preferred "selected sports programs"; "individual lectures," "formal course offerings," and "classical music" were ranked last and in that order. If the two different program classification systems are truly comparable, it would seem there is some agreement in the findings. These data seem to suggest that those viewing telecourses do so *in spite* of their preferences and not *because* of them.

Preferences for Future Telecourses. Related to the question of program preference is the item in the schedule designed to elicit information on other telecourses desired by the group. The respondents named, in this open-end question, a total of 369 subjects ranging from accounting to zoology. An analysis of these makes clear the highly specialized interest of the particular group under study. That is, courses directly related in subject matter to the subject material of the present course constituted 18.5% of those requests. Advanced algebra was named on 27 questionnaires, geometry on 26, and trigonometry was listed on 25 returns. The science subjects, chemistry, electronics, physics, and general sciences were named in 71 of the responses. Combining the figures for science and mathematics, it is found that 35.3% of all responses were in these categories. Only two subjects outside of the science field were mentioned more than ten times—these were general history, fifteen mentions, and Spanish which was noted in twelve instances. This desire for more courses in mathematics and science does not agree with the observation by Evans that the literature indicates that "liberal arts courses were preferred to science courses."[24] This would suggest that the kind of course that is requested for the future depends upon the subject of the present telecourse the student is engaged in viewing. This finding further strengthens the observation that telecourse audiences seemingly demonstrate strong interests in particular areas of learning, and not in educational programs in general.

Previous Telecourses Viewed. Continuing the above discussion, it is important to consider here the history of the individuals in terms of telecourses viewed in the past. The hypothesis suggested above, that audiences tend to be specific in the selection of telecourses they watch, is borne out by the response of the sample when questioned if they had watched a telecourse before. Of the 234 responses to the question, 151 or 64.5% were negative. To further check this response, the respondent answering "Yes" to the first question was asked to list either the subjects and/or titles of the telecourse viewed. By checking these responses against the telecourses known to have been broadcast, over either commercial or educational stations, it was possible to assess the accuracy of response. Of the 83 affirmative answers, 74 of the respondents named one or more subjects and/or titles. Of the total 146 programs so listed, 21 of them were

[23] Evans, Roney and McAdams, *op. cit.*, p. 278.

[24] Evans, Roney and McAdams, *op. cit.*, p. 278.

found to be in error. In addition to providing us with some idea of the accuracy of the recall, the relatively low number of "Yes" responses suggest again the specific interest of this group in a particular subject area. This observation should be qualified, however, with the realization that the newness of the educational television station in the Seattle area and the consequent limitation of the total number of programs available to the interested viewer may be affecting the response to this item at this time.

III. Summary and Conclusions

Summary. Two hundred fifty six viewers of the telecourse "Intermediate Algebra" were sampled by means of a mail questionnaire in order to determine certain selected demographic, background, and behavioral characteristics. Most of these viewers were men (73.9%). The viewers studied here appear to be older than those students enrolled in Evening Class program at the University of Washington. Fifty-seven and two-tenths per cent of the sample fell into the age bracket 36 to 55 and over, while only 30.1% of the Extension class sample were in this group. This finding also indicates that the group studied here was older than the general television audience. It was found that most of the viewers (56.1%) had had a similar course, although not via television, before undertaking the telecourse. When asked how they heard about the telecourse, the largest percentage (72.2%) indicate the television station and the newspaper were the sources of information. Educationally, this sample was better prepared than the general television audience. Six per cent had completed the eighth grade, 58.0% the twelfth, 28.9% college, and 6.0% some graduate work. Sixty-nine and seven-tenths per cent of the sample were employed, and 32 of these were regularly enrolled as students. Most of the respondents had some specific purpose in viewing the course, and 67.1% said they thought the course would help them in their "professional advancement." Almost half (49.6%) of the viewers reported an annual income of over $5000. Most of these questioned stated they had one set in the home, and 79.2% of those responding indicated they shared this set with from one to three persons. Total television viewing time by the sample approximated that of the general audience, and 58.1% of them said they viewed two to three hours per day. The respondents ranked "Drama" as their favorite type of program, with "Education" listed as fifth choice out of six possible. When asked to name other subjects they would like to have presented as telecourses, 423 responses were received. Of these, 35.3% requested more courses in mathematics, and science. Only two courses outside of that field were mentioned twelve or more times. This is in contradiction to the literature on the subject. Finally, relatively few of the sample had viewed previous telecourses (26.6%) but the individuals who did so were successful in correctly identifying the subject or title of the course they had watched.

Conclusions. Within the limitations of the present study two conclusions are suggested:

1. It would seem that educational television audiences, as exemplified by the sample used here, are different from the general television audience, and that these differences set them apart as distinct audience groups.
2. It would further seem that the audience viewing a particular telecourse is characterized by a strong interest in that specific academic subject, and not in educational television programs in general.

Reasoning from these conclusions, the author has been led to the consideration of two other problems which are now under study. They are:

1. A comparison of the characteristics of the audiences viewing telecourses with different academic subjects.
2. A study designed to assess more directly the attitudes of the individual, who uses a telecourse, toward television as a method of instruction.

With these studies in mind, it is suggested that the following areas are in need of further investigation:

1. Some effort should be made to determine if the sample as selected in this study is truly representative of the audience viewing an educational telecourse.
2. The need for expanded exploration of the attitudes and expectations of the telecourse viewer is clearly indicated by this present study.

AN EXPERIMENTAL STUDY OF RETENTION IN EDUCATIONAL TELEVISION*

DANIEL W. MULLIN
Tulane University

IT has been said that television is the greatest achievement in communication on a mass basis since the invention of the printing press. The textbook, a product of the press, is a convenient conveyor of educational material. Nothing is more accessible than the text. It may be acquired at a nominal cost; it may be used at any time, in any place, at various reading speeds, at the will of the student, in part or as a whole. In contrast, television is inflexible. It must be used at a scheduled time, in a prescribed place, at a pace determined by the instructor or the director, at the will of someone other than the individual student, usually in its entirety rather than in parts, selected to meet the needs of the student. And educational television is not accessible to the individual student at a nominal cost.

Therefore, since television cannot be adjusted to the individual student, it is important to determine the general conditions under which television may be utilized as a tool of education.

The purpose of this study is to determine whether home viewing or classroom viewing is more effective in promoting learning and retention, and to determine the effect of the presence or absence of motivation in home and classroom viewing. For experimental purposes, motivation was defined as an extrinsic award of monetary value which the viewer received for high learning scores measured in terms of remembered information presented solely by television. Conditions that normally prevail in a particular home were not consciously altered by the experimenter. In the classroom environment, conditions were altered only by the addition of a television set and the restriction that only twenty viewers were permitted in one classroom. Classroom viewing took place in the presence of a teacher.

Relative to the differences in experimental conditions, it was desired specifically to answer the following questions:

(1) How effective is television in teaching information to the *unmotivated* viewer in the *home*?

(2) How effective is television in teaching information to the *unmotivated* viewer in the *classroom*?

(3) How effective is *motivation* in increasing the retention of the viewer in the *home*?

(4) How effective is *motivation* in increasing the retention of the viewer in the *classroom*?

(5) Does the individual viewing *in the home* differ in retention score from the individual viewing *in the classroom group*?

It is generally accepted that intrinsic motivation, if it is strong, or some form of extrinsic motivation is effective in bringing about learning and retention.

There may be a difference between the two environments, however, which operates in the direction of making the *home* viewer more capable of learning than

*Based on Ph.D. dissertation, University of Michigan, 1956, directed by Edgar E. Willis. This research was supported by a grant-in-aid from the Louisiana Educational Television Commission administered by Tulane University.

the *classroom* viewer. Le Bon[1] theorized that the individual in an organized crowd suffers a deterioration of his capacity to act as a civilized person for two reasons: (1) his primitive emotions are stimulated, (2) his intellectual ability is lowered. There is a question, of course, as to whether or not the individual in the classroom group may be compared to Le Bon's individual in the crowd. The opinion, a common one, that differences exist is shared by Lindgren who says that, "People tend to act differently as members of groups from the way they act when they are not members of groups."[2] McCurdy and Lamber[3] conducted an experiment which, in general, led to the conclusion that any advantages of the group approach to intellectual tasks appear to be extremely elusive. Permutter,[4] however, concluded from an experiment involving group learning of nonsense syllables that there is some indication of conditional superiority of a group effort over individual effort. *Conclusive* evidence that the individual is inferior, superior, or identical to himself in the group has not been produced.

The use of television (as a formal educational tool) has been suggested as necessary to a more effective educational system. An increasing need for competent teachers and the shortage of classroom space are two of the major problems that face our educators. Can television aid in solving these problems?

Western Reserve University has experimented with a program of formal college-credit courses taught by television exclusively. John F. White, Vice-President in Charge of Development, reported that, "As a result of our experiment, we do know that formal college-credit courses can be presented successfully over television. We know, too, that the students who enroll in such courses do as well and learn as much as the typical student in the classroom."[5] Husband,[6] using the same instructor in the classroom and on television, obtained results that suggested that higher achievement occurred in the television-instructed group; however, statistically significant differences were not obtained. Rock, Duva, and Murray[7] tested the effectiveness of television, television recordings, and conventional classroom instruction. Comparisons indicated that television instruction is as good or better than conventional classroom instruction.

One question that occurs with respect to these experiments is the possible stimulation of interest developed by the novelty of television instruction, a novelty that does not exist in a class taught by conventional classroom techniques. To what extent did the novelty of being taught by television stimulate the interest and retention of the subjects?

For example, Jackson[8] conducted an experiment in which matched groups

[1] Gustave Le Bon, *The Crowd* (London, 1914), pp. 36-37.

[2] Henry Clay Lindgren, "The Effect of the Group on the Behavior of the Individual," *Education*, LXXIII (February, 1953), 383-387.

[3] Howard G. McCurdy and Wallace E. Lamber, "The Efficiency of Small Human Groups in the Solution of Problems Requiring Genuine Co-operation," *Journal of Personality*, XX (June, 1952), 478-493.

[4] Howard V. Permutter and Germaine De Montmollin, "Group Learning of Nonsense Syllables," *Journal of Abnormal and Social Psychology*, XLVII (October, 1952), 769.

[5] John F. White, "TV or Not TV?" *National Educational Association Journal*, XLII (January, 1953), 34.

[6] R. W. Husband, "Television Versus Classroom for Learning General Psychology," *American Psychologist*, IX (1954), 181-183.

[7] Robert T. Rock, James S. Duva, and John E. Murray, *Training by Television, The Comparative Effectiveness of Instruction by Television, Television Recordings, and Conventional Classroom Procedures*. A Report Prepared by the Special Devices Center, Office of Naval Research, Department of the Navy, Port Washington, L. I., N. Y. (November, 1951), pp. 3-9.

[8] Robert Jackson, *Learning From Kinescopes and Films*. A Report Prepared by the Special Devices Center, Office of Naval Research, Department of the Navy, Port Washington, L. I., N. Y. (April, 1952), pp. 11-14.

viewed the same kinescope. One group was told that they were looking at a kinescope, an uncommon experience. The other group was told that they were looking at a training film, a common experience. A second pair of matched groups viewed the same training film following identical viewing instructions. In both cases, the learning score of the subjects who were told that they were viewing a kinescope was significantly higher than that of the subjects who were told that they were viewing a training film. The results of the experiment suggest the probability that a new educational tool is more effective than the conventional tool, *simply because it is new and unusual.* This *novelty* effect would probably be reduced by continued usage or practice in viewing.

An experiment conducted by VanderMeer,[9] however, indicates that the student who has had practice in learning from film is superior to the student who has had no practice when both students are subsequently taught by the same film method. Thus it would appear that a decrease in learning due to a loss of the *novelty* effect may be balanced by an increment in learning due to *practice* in viewing. Whether or not these results are applicable to a comparison between television and classroom is a question subject to further research. The restrictions inherent in the comparison of educational television with another medium such as conventional classroom instruction are absent in an experimental design involving the comparison of television with itself as is the case with this study.

[9] A. W. VanderMeer, *Effect of Film-Viewing Practice on Learning from Instructional Films.* A Report Prepared by the Special Devices Center, Office of Naval Research, Department of the Navy, Port Washington, L. I., N. Y. (November, 1951), pp. 3-9.

I. Procedure

A twenty-three minute educational television program was produced according to as many of the tested principles of television instruction as could be followed within the framework of experimental conditions. "Explorations in Space" was chosen as the title of the program which presented information primarily concerned with the nine planets in our solar system, the nearest universe, and a few of the problems man faces in leaving earth for space travel. This subject is highly adapted to instruction by television in that the use of visual aids is a natural and essential procedure in the conventional classroom presentation of the material.

Dr. Joseph F. Thomson, the instructor, had had experience both in classroom teaching procedures and in performance techniques before the television camera. He is a member of the faculty of Tulane University and was near the end of his second year of television experience on his own daily weather program at the time of this special educational television broadcast.

All of the students participated in the experiment on a volunteer basis. The subjects were members of both sexes at the eleventh grade level in three high schools in the same city. They were divided into four experimental groups and two control groups.

Group 1 included subjects who were not motivated to learn and retain. That is, they were instructed to judge the quality of the presentation, but they did not know that they would be tested. *They viewed as individuals in the home.*

Group 2 was composed of subjects who were motivated to learn and retain. They were informed that they would be tested and rewarded according to the learning scores achieved. *They viewed as individuals in the home.*

Group 3 was composed of subjects who were not motivated. *They viewed as members of an in-school group.* The viewing took place in a classroom in the presence of a proctor.

Group 4 was composed of subjects who were motivated. They were informed of the post-viewing examination and the awards which would be received by those who attained high learning scores. *They viewed in a classroom in the presence of a proctor.*

The subjects in *Group 5*, a control group of casual viewers, were not aware of their role in an educational television experiment before or during the telecast. The only pre-telecast contact was made in the form of a publicity announcement of the material to be treated on the program. *These subjects viewed as individuals in the home.*

The subjects in *Group 6*, a control group, were unaware of their role in an experiment before or during the telecast. *These subjects did not see the program.* The only pre-telecast contact was made in the form of the same publicity announcement given to *Group 5*.

At the beginning of the first post-viewing session, all of the subjects were urged with identical instructions to do as well as they could on the immediate recall test. All of the subjects were informed of the awards to be received by those who obtained high learning scores. Following the immediate recall test, the members of all groups were informed that the retention awards would be announced at another session two weeks later. Two weeks later, at the second post-viewing meeting, all of the subjects were invited to take essentially the same test again for identical or additional awards. Following this delayed recall test, the awards on the immediate recall test were announced. Each subject had competed only with members of his viewing group. There were six sets of identical awards for each of the test administrations. The two sets of examinations, composed of the same fifty multiple-choice items, given in two random forms, were graded. The raw scores for the subjects plus their face-sheet information were used as the basis for the analysis.

Face-sheet information was derived in part from each subject's answers to questions contained in the application form filled out by each volunteer student. The question, "Are you interested in *space travel?*" was designed to provide an idea of the subject's attitude toward the program material. Although general science is a required course in the junior high school, not all of the senior high students had had the course. Therefore, the question, "Have you completed one course in general science?" was included in order to gain some idea of the student's background in this area. Two questions were asked to provide a basis for determining the socio-economic index for each subject: (1) What does your father do for a living? (2) Is your father self-employed? The index resulting from the preceding questions was supplemented by comparing the subject's home address with the 1950 census tracts to identify the socio-economic area in which each subject lived. Other items of face-sheet information included the age, the sex, and the IQ of each subject.

The motivated subjects who indicated that they were willing and able to take part in the experiment totaled 120. The unmotivated subjects who volunteered numbered 96. The number of subjects who participated in the experiment decreased from session to session. Fortunately, only comparatively few subjects were necessary to measure significant differences between the mean scores of the groups at the .01 level of confidence. Volunteer subjects seem to require a high level of extrinsic motivation if

TABLE I
MEAN TEST SCORES

Group	Immediate Recall N	Mean	SD	(Subgroup) N	Mean	SD	Delayed Recall N	Mean	SD
1	17	22.65	4.69	11	20.82	4.02	11	24.18	7.73
2	11	32.09	9.03	7	36.43	8.52	7	34.86	9.08
3	13	24.46	6.97	8	25.75	7.85	8	28.50	7.37
4	10	27.50	6.35	6	29.17	7.68	6	27.00	9.08
5	8	29.63	10.95	4	36.75	2.87	4	36.50	3.42
6	19	15.42	8.17	12	16.67	9.51	12	19.42	7.99
Total	78	23.86	9.32	48	25.25	10.25	48	26.65	9.46

they are expected to remain with the project throughout all of the sessions. The question may arise, then, as to the experimenter's liberty in referring to some of them as unmotivated. A distinction was made between those subjects who were requested to judge (unmotivated) the quality of the presentation and those who were urged to learn and retain the information (motivated).

II. RESULTS

The means and standard deviations for each of the six primary groups for both the first and second administrations of the test are presented in Table I. In addition to the means for the entire group on the immediate recall test, separate means are presented for the subgroup that was tested after the two-week interval. The subgroup consisted of those subjects who took both the immediate and the delayed recall tests.

A comparison of the mean score for *Group 5*, control subjects, with the mean scores of the 4 experimental groups would seem to indicate that the *casual* viewer in the home has achieved a higher learning score than all other subjects with the exception of those in *Group 2*, motivated in the home. As a control group, *Group 5* achieved a mean score that is most probably unreliable because of the following conditions: (1) some of the subjects in *Group 5* received their immediate recall test two days late; (2) the four subjects in the subgroup had Otis Mental Ability Test scores of 111, 115, 132, and 133; the mean IQ scores of the four experimental groups ranged from 100 to 109.

The differences among the means of all groups, based on all of the subjects who were tested during the first administration, immediate recall, are significant at the .01 level of confidence when evaluated in a simple analysis of variance design. These results are presented in Table II.

TABLE II
ANALYSIS OF VARIANCE TEST SCORES
PRIMARY GROUP

Source of Variation	df	Mean Square	F
Between groups	5	508.28	8.81**
Within groups	72	57.71	
Total	77	....	...

**Significant at the 1% level.

Bartlett's test for homogeneity of variance was used to substantiate the analysis of variance findings. A non-significant Chi-square of 9.53 was obtained and indicates that the variance for the six groups are not heterogeneous on this measure. This leads to the conclusion that the significant F-test obtained in the analysis of variance is unequivocally due to differences in group means. The differences in the means indicate that the groups ranged from high to low mean scores as follows:

Group 2, motivated in the home

Group 4, motivated in the school

Group 3, unmotivated in the school

Group 1, unmotivated in the home

Group 6, non-viewer

Group 5, casual viewer in the home, achieved a mean score higher than that of *Group 4* and lower than that of *Group 2,* but the results are not included in the above order for the reasons previously stated.

A similar analysis was performed for the results of the first administration using the data for the subjects who also took the second test—the subgroup. Similar to the results in Table II, these results are presented in Table III.

TABLE III
ANALYSIS OF VARIANCE OF TEST SCORES
FIRST ADMINISTRATION SUBGROUP

Source of Variation	df	Mean Square	F
Between groups	5	519.58	9.31**
Within groups	42	55.79	
Total	47	....	...

**Significant at the 1% level.

The differences among the means of the six groups for the second administration of the test, *delayed recall,* were also evaluated with analysis of variance techniques. The results of this analysis are presented in Table IV.

TABLE IV
ANALYSIS OF VARIANCE OF TEST SCORES
SECOND ADMINISTRATION

Source of Variation	df	Mean Square	F
Between groups	5	316.51	5.07**
Within groups	42	62.44	
Total	47	....	...

**Significant at the 1% level.

Bartlett's test for homogeneity of variance resulted in a Chi-square of 3.14 which is not significant. The F-test in Table IV may then be used as a basis for rejecting at the .01 level of confidence, the hypothesis of no difference among the means of the six groups on the second administration.

Having demonstrated that differences exist among the groups for both the first and the second administrations, the relationship between the results obtained from these two administrations was then investigated. It was concluded that the mean differences for the six groups on the second administration were in relation to the first administration differences and may be explained by the measurements obtained on the first administration.

The data from the first administration for *Groups 1, 2, 3,* and *4, the experimental groups,* were analyzed in a factorial design to test for the effects of environment, motivation, and possible interaction effects. The analysis is complicated by the presence of disproportionate subclass (sex, expressed interest, etc.) members. The completed analysis of variance is presented in Table V.

TABLE V
COMPLETED ANALYSIS OF VARIANCE

Source of Variation	df	Mean Square	F
Environment	1	3.17	
Motivation	1	516.32	11.49**
Interaction	1	145.64	3.24
Individuals	47	44.95	

**Significant at the 1% level.

The unbiased estimate for the mean difference between home and school environment is .25 and the unbiased estimate for the mean difference between the motivated and the unmotivated groups is 7.38. This last difference is significant at the .01 level of confidence while the interaction and environment effects are negligible.

A similar analysis was undertaken for the data obtained in the second administration. The presence of interaction effects between motivation and the en-

vironment, however, was indicated in that the unmotivated subjects achieved a higher mean score in school whereas the motivated subjects achieved a higher mean score at home. The results of the analysis indicated that neither of the experimental variables, motivation and environment, produce significant effects. The interaction effects approach significance at the .05 level of confidence, but little interest attaches to this fact in the absence of differences in the main effects.

Additional information concerning several characteristics of the six groups was obtained from the face-sheets. The correlation between the IQ scores and scores on the first administration of the test for *Groups 1, 2, 3, 4, the experimental groups,* may be regarded as chance fluctuation that could have arisen in a population where the true correlation between IQ score and test score is zero. The correlation obtained with *Group 5* is significantly different from a zero correlation at the .05 level of confidence, while the correlation obtained with *Group 6,* non-viewer, is significant at the .01 level of confidence.

Other characteristics of the six groups are not subject to evaluation in terms of differences between the various groups due to low numbers within each subclass. Inasmuch as they might be variables operating in the present study, it is unfortunate that they were not subject to control in the experimental design.

There is some indication in the present population that one could control for differences in interest and enrollment in science rather well merely by equating groups with regards to the proportion of male and female members. Tests for the statistical independence of the sex variable indicated a positive relationship between sex and enrollment in a science course and between sex and interest expressed in the lecture material. In lieu of an evaluation of the differences in the distribution of those characteristics in the various groups (numbers in subclasses are not large enough for the evaluation), it seemed feasible to determine whether or not there appears to be a relationship between the sex variable and the mean scores within the experimental groups. A test for differences between the male means and the female means within the experimental *Groups 1, 2, 3,* and *4* indicated that there are no significant differences at the .01 level of confidence. It is interesting to note that the difference between the means in *Group 6,* non-viewer, is significant at the .01 level of confidence. Apparently, the males among the non-viewers were either better informed or better at guessing the correct answers to the questions on space travel and the planets. The means among the females are consistently lower than the means among the males. However, failing to establish significant differences between the male mean and the female mean in each experimental group would seem to indicate that the females learned more or retained more than the males. And there is a probability that they had *more* to learn about space travel and the planets than did the *boys.* A reversal in this relationship would take place, no doubt, if the subject had been one of particular appeal to females.

A disadvantage of the volunteer method was apparent in the rapid decrease in the number of subjects from session to session during the experiment. The remaining subjects must be defined as students who are most highly co-operative or highly confident of attaining awards on a competitive basis. They are also of above average intelligence and they tend to concentrate in income brackets that can be described as medium to high. These are the stu-

dents to whom the results of this experiment apply directly. There is, however, no apparent reason for assuming that other than relative differences would have occurred if average or below average students had been tested under the same conditions. In a relative sense, the conclusions of the experiment may apply generally.

III. Conclusions

The statistical results of the experiment indicate that the questions posed by the study may be answered as follows:

Question: How effective is television in teaching information to the *unmotivated* viewer in the *home? Answer:* A significant difference between the mean score of the non-viewer and and the mean score of the unmotivated viewer in the home indicates that some learning does take place. When compared with the mean score of the motivated viewer, however, it is apparent that the efficiency of learning is not comparable to that of the motivated viewer.

Question: How effective is television in teaching information to the *unmotivated* viewer in the *classroom? Answer:* Numerical differences between the means suggest that the classroom environment is more conducive to learning and retention than is the home environment in the case of the unmotivated viewer.

Question: How effective is *motivation* in increasing the retention score of the viewer in the *home? Answer:* Relative to the learning score of the unmotivated viewer, the learning score of the motivated viewer is significantly higher.

Question: How effective is *motivation* in increasing the retention score of the viewer in the *classroom? Answer:* Relative to the unmotivated viewer in the classroom, the motivated viewer learns significantly more. The analysis indicates quite conclusively that the increment in learning is due to motivation.

Question: Does the individual viewing *in the home* differ in retention score from the individual viewing *in the classroom group? Answer:* The difference between the mean scores of the two groups of unmotivated viewers suggests that the school environment is more conducive to unmotivated learning. The difference between the means of the motivated groups, however, suggests a superiority of the home environment over the school environment in educational television viewing.

Neither the home nor the school environment is consistent in its effect, if any, on retention under different conditions of motivation. In fact, results indicate contradictory effects in that the unmotivated subject achieves a higher score in school whereas the motivated subject receives a higher score in the home.

The results of the study indicate that the effects of motivation and environment on retention may be summarized in the following conclusions:

(1) Motivation is significantly effective in increasing the retention scores of both the home viewer and the classroom viewer.

(2) Neither of the two environments considered is consistent in its apparent effect on the retention scores of both the motivated and the unmotivated viewers.

Inasmuch as the mean scores suggest that the motivated student may learn more at home than he does in the classroom, educational television in the home may be effective as part of the solution to the increasing shortages of teachers and classrooms.

AN EXPERIMENTAL STUDY OF THE ACQUISITION OF INFORMATION FROM THREE TYPES OF RECORDED TELEVISION PRESENTATIONS*

JOHN HOLWAY ULRICH
University of Pittsburgh

I. Statement of the Problem

MANY educators today are hopefully looking toward television for aid in some of the problems that are manifest in all areas of education. Though films and television are becoming increasingly important in education, insufficient research under adequately controlled conditions has been done on these media. This kind of research is necessary to supply information needed to use these media to maximal advantage. Specifically, this investigation was designed to determine whether eighth grade pupils retained more information from observing a kinescope recording[1] of a lecture without visual aids, one with aids handled by the lecturer, or the same lecture with aids that were merely flashed on the screen.

Completed research indicates that television can be used as a teaching medium. Kanner, Runyon and Desiderato[2] in a study conducted under the auspices of the Human Resources Research Office (HumRRO) at Camp Gordon, Georgia in 1953, found that after carefully matching conditions between television instruction and regular instruction methods, television instruction was at least as effective as classroom teaching and more effective in the case of low aptitude groups. The materials for instruction were taken from the eight week Army Basic Training program which included both content and performance materials. The criteria of effectiveness were pretests, immediate tests (administered after each hour of instruction) and retention tests which measured the amounts of information retained by the trainees 30 days after original instruction.

It was also found that retention was as good for television instructed groups as the regular classroom instructed groups. It was further noted that instruction by kinescope recording was also as effective as classroom instruction. Similar conclusions are drawn in a study conducted by Rock, Duva and Murray[3] in which television was frequently found to be superior to local instruction when used to instruct groups of Naval Air Reservists. Tests that measured the acquisition of information were used as the effectiveness criteria.

After a relatively exhaustive survey of television utilization in training methods, Fritz, *et al.*[4] observed that

*Based upon Ph.D. dissertation, State University of Iowa, 1955, directed by H. Clay Harshbarger. The writer was also assisted by Sam L. Becker.

[1] Kinescope recordings are 16 mm. sound film recordings of a television production. They are made with a device called a kinescope recorder which allows a motion picture to be recorded from a special television receiver tube.

[2] Joseph H. Kanner, Richard P. Runyon, and Otello Desiderato, "Television in Army Training: Evaluation of Television in Army Basic Training," *HumRRO Technical Report 14.*

[3] Robert T. Rock, Jr., James S. Duva, and John E. Murray, "The Effectiveness of Television Instruction in Training Naval Air Reservists," *Technical Report SDC 476-02-52,* Special Devices Center, Port Washington, Long Island, New York.

[4] Martin F. Fritz, "Survey of Television Utilization in Army Training," *Human Engineering Report SDC 530-01-1,* p.5, Special Devices Center, Port Washington, Long Island, New York.

"Television can be advantageously integrated into the Army training program." When comparing learning effected through the use of kinescopes and film, Jackson[5] noted that trainees tended to learn more when they *thought* they were viewing kinescope recordings as opposed to standard films. In another study by Rock, Duva and Murray[6] it was found that television was not only effective in training large groups of Army Field Force Reservists but that motivation and retention were both high, even after eight weekly sessions. These sessions consisted of viewing programs that were concerned with the operation of an Army Division in combat. The viewers were reservists assembled in 160 separate groups and each reservist was tested after each viewing session.

From the above briefly discussed research projects two conclusions pertinent to this study emerge:

1. Television *can* be used as a teaching device.
2. Live television and recordings of live programs (kinescope recordings) are about equally effective when used in the classroom.

The producer of educational television programs has a number of presentation methods available to him. One of the most frequently used is the lecture or talk with or without visual aids. This study was concerned with such lectures. Specifically, this investigation was designed to determine whether eighth grade pupils might retain more information by observing a kinescope recording of a straight lecture when using no visual aid, the same lecture using visual aids handled by a lecturer, or the same lecture with visual aids flashed on the screen. In this report the straight lecture will be called *The Straight Verbal Lecture,* the lecture with visual aids handled by the instructor will be called *The Instructor Facilitated Aids Lecture,* and the lecture where the aids are flashed on the screen will be called *the Camera Facilitated Aids Lecture.*

II. Procedure

The food resources of Africa served as the lecture topic. This subject was chosen because it is seldom (if ever) included in courses taught at or before the eighth grade level. A general body of information concerning the food resources of Africa was collected from one major source and a seventy-six item (multiple choice-four response and true-false) test was constructed from this information. From this test were to come the items that would guide the construction of the experimental lectures. It was deemed necessary to determine which test items covered information not readily known by the sample population. The following procedure was followed:

1. A group of five adults for five weeks studied the specific information that went into the test. They were called "The Board of Experts."
2. At the end of five weeks the seventy-six item test was administered to this "Board of Experts."
3. Three elementary schools were selected from the Chicago School system. (The Chicago Schools were selected to supply the sample for this project.) The three schools were selected as representative of the Chicago Schools in general.
4. The seventy-six item test was administered in one eighth grade class in each of the schools. A total of 114 pupils completed the test.
5. In order to select those items that appeared to discriminate well between those who knew the materials and those who did not, the

[5] Robert Jackson, "Learning from Kinescopes and Films," *Technical Report SDC 20-TV-1,* Special Devices Center, Port Washington, Long Island, New York.

[6] Robert T. Rock, Jr., James S. Duva, John E. Murray, "A Study of Learning and Retention from Television Instruction Transmitted to Army Field Force Reservists," *Technical Report SDC 476-02-S3,* Special Devices Center, Port Washington, Long Island, New York.

following selection procedure was followed:

a. Those multiple choice items that were answered correctly by 100% of the "Board of Experts" and by 40% or less of the 114 elementary pupils were retained.
b. Those true-false items that were answered correctly by 100% of the "Board of Experts" and by 50% or less of the 114 pupils were retained.

A total of nineteen items (17 multiple choice and 2 true-false) were retained and used as a core for the experimental lecture. Because the experimental variable was in the presentation method rather than in the verbal form of the lecture, one lecture was devised to fit the three presentation methods. The facts contained in the nineteen test items were worked into a fifteen-minute lecture which was equally adaptable to all three presentation methods. It was then converted into three television scripts; the straight verbal lecture script, the instructor facilitated aids lecture script and the camera facilitated aids lecture script. For each of the nineteen items there was a corresponding unit of information in the script. Each one of these units was carefully considered to avoid undue emphasis upon any particular area. An attempt was also made to delete any extraneous materials from the scripts.

Poster type visual aids were planned for each unit of information. The following five classes of visual aids were to be developed:

1. Enlarged photographs.
2. Outline maps with particular geographic information (such as the location of a city) printed on the maps.
3. Outline maps with cartoon drawings to indicate some particular activity or special geographic location of some activity.
4. Graphs and/or charts with printed information.
5. Pictograms that were similar to the items in 4 above except that the information supplied was supplemented by drawings or cartoons.

Following the completion of the scripts, they were submitted to ten judges, all of whom were State University of Iowa faculty members. The judges evaluated the three scripts in terms of suitability of content, language, age level, and general lecture quality. The judges made their evaluations on a specially prepared form and also in personal conversation with the researcher. After the evaluations were made, the author wrote the final drafts of the scripts incorporating the suggestions of the judges.

After careful examination of the informational units, visual aids were constructed. In all but two cases one visual aid was constructed to correspond to each test item. In one case an aid covered two test items, and, in the other, a test item was covered by two aids. The aids were produced on colored poster card in the 3 by 4 aspect ratio required by television.

The lecturer chosen for this project was selected because of his broad experience in front of television cameras and his experience in working with the medium both experimentally and artistically. His appearance and maturity of manner were judged as excellent for this undertaking. Dr. James E. Lynch, the lecturer, prepared his presentations under the direction of the author. The script was committed to memory, but to assure identical presentation in all three treatments, cue cards were held near the camera at which the lecturer was looking. In this way, verbatim repetitions of the lecture were given in all three treatments. The following describes the nature of each lecture treatment as it was delivered for the kinescope recording:

1. TREATMENT I (Straight Verbal Lecture): The instructor presented the body of information in a direct manner to the cameras. He used normal

gestures and employed no visual devices.

2. TREATMENT II (Instructor-facilitated aids lecture): Visual devices were used by the speaker as he delivered the same lecture.
3. TREATMENT III (Camera-facilitated aids lecture): The lecturer delivered the same speech as in Treatments I and II, the delivery being identical to that in Treatment I. The visual devices were switched on the screen when appropriate to the material and always in a close-up aspect. At no time were the lecturer and a visual device seen together on the screen.

The technical quality of the three kinescope recordings was equally satisfactory. Images were clear when projected over standard 16 mm. projection systems. The sound track was easily understood at all times.

To measure the relative retention of material from the presentation methods, a random sample of forty eighth-grade classes was selected from a sixty school cross section of the city school system of Chicago, Illinois. Each of these schools was randomly assigned by a table of random numbers to one of four groups: Treatment I, Treatment II, Treatment III and the control group. It would have been desirable to assign randomly the schools to a schedule of testing; however, expedience ruled this out. Hence, a map of the city of Chicago was examined and the schools were divided into proximity groups, each of which could be tested in a single day. Whenever practicable, adjustments were made in scheduling to avoid too great a concentration of any one treatment at any particular time of the day or week.

The kinescope recordings were projected on a model 285-Bi-30 Bell and Howell sound projector. The method of presentation of materials was standardized for all treatments. Pupil preparation was identical in each case except for the control, where the wording had to be altered to exclude the information about the kinescope recording.

Immediately following the presentation of the kinescope recording, *Test A* was administered. *Test A* was composed of the nineteen test items that had been randomly ordered in the testing instrument. International Business Machine answer sheets were employed in the testing procedure. Following the completion of the test, each class was informed that the author would return in one month. They were not told that they would be tested again, and there was nothing said about the materials on which they had been tested. The teachers were not told the correct answers to the test, though all teachers were allowed to see the questions. No materials were left with the teachers or the pupils. Extreme care was taken at all times to treat each class, teacher, and principal in an identical manner.

The retest (Test B), was the same as Test A, except for a rearrangement of the item order (by random assignment through the use of a table of random numbers). *Test B* was administered to the classes one month after *Test A* had been given. The testing procedure was similar except that more schools were tested per day since there were no kinescope recordings to be screened for the second test. The method of testing was standardized for all the classes. Again IBM scoring sheets were used to collect the test data.

III. ANALYSIS OF THE DATA

For the initial analysis of the data, a groups-within-treatment design was utilized, with the class mean becoming the criterion measure. With this analysis of variance procedure, the 5% level of

TABLE I
MEAN SCORES FOR TEST A

	Control	Treatment I Straight Verbal Lecture	Treatment II Instructor Facilitated Aids Lecture	Treatment III Camera Facilitated Aids Lecture
Class Means	5.089	13.556	13.736	15.975
	5.889	11.553	13.275	13.976
	5.857	14.171	15.333	13.800
	4.021	15.341	12.923	14.184
	4.500	8.268	14.222	11.441
	4.944	9.950	13.057	11.500
	4.243	9.150	13.200	12.757
	4.225	8.450	14.976	10.417
	5.000	9.500	13.971	13.029
	4.432	14.327	15.258	15.821
Group Means	4.920	11.427	13.995	13.290

confidence was set at which to reject the hypothesis of no difference. In all cases where an analysis of variance proved significant, t-tests were applied to each individual pair of treatment means.

IV. RESULTS

Test A

A total of 1,518 pupils were tested in the *Test A* sample population. Of this total 404 were in the control group, 368 in Treatment I group, 387 in Treatment II group, and 359 in Treatment III group. Class mean scores and group mean scores may be examined in Table I. These mean scores seem to indicate that a considerable amount of information was gained by each of the treatment groups and that little or no knowledge of the subject was observed in the control group. It may also be noted that the two mean scores which were most similar (13.995 and 13.290) were from the treatment groups which saw the kinescope recordings containing visual aids.

In both *Test A* and *Test B* two analyses of variance were employed. The first included the control group, the second did not. This was done to test most rigorously for differences between the treatment groups under all conditions. However, for the purpose of this report those analyses that tested the treatment groups only and excluded the control groups will be reported. The F's obtained for those analyses that included the control groups were much higher than those obtained without, and in both cases of *Tests A* and *B*, significant at the pre-set 5% level of confidence. The primary reason for including a control group was to determine whether there was any statistical evidence of learning per se in any of the treatment groups. The extremely large F that resulted left little doubt of this.

Table II describes the analysis of variance results when only the three experimental treatment groups of *Test A* were examined. With 2 and 27 degrees of freedom an F of 3.35 must be obtained to be significant at the 5 per cent level of confidence. Therefore with an ob-

TABLE II
SUMMARY TABLE FOR TEST A ANALYSIS OF VARIANCE WITH THREE GROUPS

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Treatments	2	35.240	17.620
Groups-within-treatments	27	103.718	3.841
Total	29	138.961	
	$F = 4.59$		

Significant at the 5 per cent level of confidence. (3.35 necessary for significance at the 5 per cent level of confidence.)

tained F of 4.59, the hypothesis of no difference between the means was rejected. This analysis indicates that statistically significant differences did exist among the three groups. It does not indicate whether such differences existed between the individual groups. For this purpose, t-tests were employed. The results may be noted in Table III. The

TABLE III
TABLE OF SIGNIFICANT DIFFERENCES BETWEEN THREE GROUP MEANS ON TEST A

	I	II
II	2.568*	
III	1.863*	.705

*Significant at the 5 per cent level of confidence. (1.799 necessary for significance at the 5 per cent level of confidence.)

critical difference was found to be 1.799, which means that the difference between any two means in these treatment groups needed to exceed 1.799 to be signficant at the 5 per cent level of confidence. Examination of Table III will indicate that there were significant differences when a treatment that involved the use of visual aids was compared to the treatment that did not. When the treatments that involved visual aids were compared, there was no statistically significant difference.

Test B

Test B was the same as *Test A* except that the order of the test items had been randomly scrambled. It was administered to the sample approximately one month later.

Table IV shows the mean test scores for each class and for each treatment. The *Test B* sample was 133 subjects less than the *Test A* sample. This loss was due to the absence of subjects during the second test or their school transfer between the testing sessions. The group means recorded at the bottom of the table seem to indicate that a considerable amount of information was retained by the treatment groups between the presentation of the kinescope recordings and the administration of *Test B*.

When *Test B* data were subjected to an analysis of variance that included the Control Group a significant F was obtained. When subsequent *t-tests* were employed it was observed that differences between the control group and each of the treatment groups and between Treatment Groups I and II were statistically significant. When the data were analyzed excluding the control groups, however, statistically significant differences among the three treatment groups disappeared. Table V summarizes the results of this analysis of variance. With 2 and 27

TABLE IV
MEAN TEST SCORES FOR TEST B

	Control	Treatment I Straight Verbal Lecture	Treatment II Instructor Facilitated Aids Lecture	Treatment III Camera Facilitated Aids Lecture
Class Means	4.970	12.710	10.667	9.387
	4.528	12.400	10.371	6.765
	3.417	5.948	11.750	8.208
	4.696	6.946	9.480	8.273
	4.649	6.588	11.219	13.886
	4.653	6.559	10.621	13.389
	4.417	10.077	12.400	10.258
	5.385	14.140	10.502	10.800
	5.444	8.944	12.971	11.278
	4.061	8.050	10.263	10.964
Group Means	4.622	9.236	11.024	10.321

degrees of freedom an F of 3.35 must be obtained for significance at the 5 per cent level. Because no significant F was obtained in the analysis of variance recorded in Table V, the t-test could not be legitimately applied.

TABLE V
SUMMARY TABLE FOR TEST B ANALYSIS OF VARIANCE WITH THREE GROUPS

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Treatments	2	16.434	8.217
Groups-within-treatments	27	134.625	4.986
Total	29	151.059	
	F = 1.65		

Not significant at the 5 per cent level of confidence. (3.35 necessary for significance at the 5 per cent level of confidence.)

Analysis of the above results seems to indicate that any difference in effectiveness between the presentation methods had a tendency to disappear in one month. Treatments II and III appear to be more effective than Treatment I in facilitating the acquisition of information for immediate recall, but they do not appear to effect any statistically significant difference in retention after a lapse of one month.

V. CONCLUSIONS

The following conclusions seem warranted:

1. Eighth grade pupils immediately recall more information from a kinescope recording of a lecture on economic geography supported by poster type visual aids than from one without such visual aids.
2. Whether the visual aids are handled by the lecturer or simply appear at the appropriate time on the screen seems to have little effect on the recall of the information.
3. Differences between lectures employing visual aids and those not employing them tend to disappear after thirty days.

Results of previous studies enable the author to make certain speculations as to the similarity in the use of kinescope recordings projected over 16 mm. film systems to actual television receiver viewing. It seems only reasonable to assume that the results of this study may contribute meaningful suggestions or create important questions with respect to educational television methods in general.

Perhaps the most far reaching implication of this study grows out of the lack of significant differences between the presentation methods at the expiration of a thirty day period. It stands to reason that a great amount of research should be undertaken to confirm or deny this phenomenon with respect to other age groups, different materials, and various objectives, especially as these conclusions could affect other media of communication and certainly other lecture or educational methods.

With the broad acceptance of the use of visual aids in education today, the results of this experiment seem to indicate a need for much evaluation of technique and practice in this field. Certainly a broad denial of the values of visual aids is not implied here, but if certain types or methods of use are ineffectual, a great saving of time, expense and manpower could be effected. It is also interesting to speculate that certain qualities in television transmitted materials may preclude the necessity for visual devices constructed to reinforce verbally presented materials. Expanded research projects along similar designs to the one here reported, may help to answer these questions. Solutions to problems of similar nature to this one also may contribute to the improvement of the educational use of the various media of communication.

ATTITUDES OF SPEECH DEFECTIVES TOWARD HUMOR BASED ON SPEECH DEFECTS*

LEOLA SCHAPER HOROWITZ
Adelphi College

MANY studies have been made on humor, and many aspects of humor are recognized to have psychological significance. Its ramifications in the need systems of individuals have certainly not been extensively explored, but we do know that humor is a psychological mechanism that satisfies some need.

Humor is certainly deeply ingrained in all of us. More important, however, is the fact that humor can be a weapon —a shaft that can inflict psychological wounds. This assumes tremendous importance when it is known that the handicapped have borne the brunt of a large proportion of our jokes. Since humor is so large a factor in our lives, and since it is often misused by the well-intentioned, as well as by the naive, it is clear that a study of attitudes toward humor involving handicapped individuals is appropriate.

The present study centers on a large and important segment of the handicapped population, speech defectives, and was carried out with two purposes in mind:

1. To determine the attitude of the speech defective toward humor based on handicaps in general and defective speech in particular.
2. To determine the attitude of the non-handicapped person in regard to humor based on handicaps in general and on speech defects in particular.

The information obtained in this study should increase our understanding of all handicapped people. It should be especially helpful to the speech clinician both in knowing the attitudes of the speech defective and in knowing the attitudes of the population in general toward those who have defective speech.

*Based upon Ph.D. dissertation, Stanford University, 1954, directed by Virgil A. Anderson.

I. Subjects and Procedure

Twenty-five subjects in each of the following groups were used—articulatory, stuttering, and control. The three groups were of comparable age range—18 to 26 years—and were exposed to approximately comparable education. Careful medical examination showed that all subjects were free from visual, auditory, and orthopedic defects. In addition, the subjects in the control group differed from the subjects in the articulatory and stuttering groups by being free from speech defects.

The subjects were told that they were participating in a study of humor. If they asked further about it, they were told that the examiner wanted to see what type of humor was preferred by college students (or adults). Each subject, alone with the examiner, rated forty-five jokes for humor on a rating scale devised by the examiner. Of the forty-five jokes there were fifteen in each of the following categories: (1) those involving speech handicaps—of these fifteen there were five each on lisping, sound substitutions other than those involving the sibilants, and stuttering; (2) those involving physical handicaps—of these fifteen there were five each on visual defects, hearing defects, and orthopedic defects; (3) innocuous jokes—those in which the handicapped were not mentioned in the joke. The jokes were randomized so that the true object

of the study might not be easily guessed.

After the jokes were rated, carbon copies of the same forty-five jokes were given to the subject. On these carbon copies the butt of the joke or the type of handicapped individual featured in the joke, if different from the butt of the joke, was underlined. The instructions were to mark the scale at any point which best described how humiliated the character underlined in the joke might be if this were a real life situation. Further instructions were given to circle the check-mark if the subject would not tell this joke to a group in which a person was present who was like the character underlined.

After the jokes were rated for the degree of humiliation involved, each subject was asked to rank the speech and physical handicaps used in the jokes in order from the most handicapping to the least handicapping for three situations—social, school, and occupational (for professional people and laborers). Each of the subjects was also asked whether he knew anyone who had any of the defects and whether he considered himself to have one of them. An interview was then held with the subject to determine the meanings of the ratings he had assigned and also to clear up certain points which a rating scale could not reveal adequately.

II. Analysis of Ratings

1. The three groups of subjects do not differ significantly from each other in their ratings on the overall humor of the jokes (See Figure 1 for results 1, 2, 3, 4).

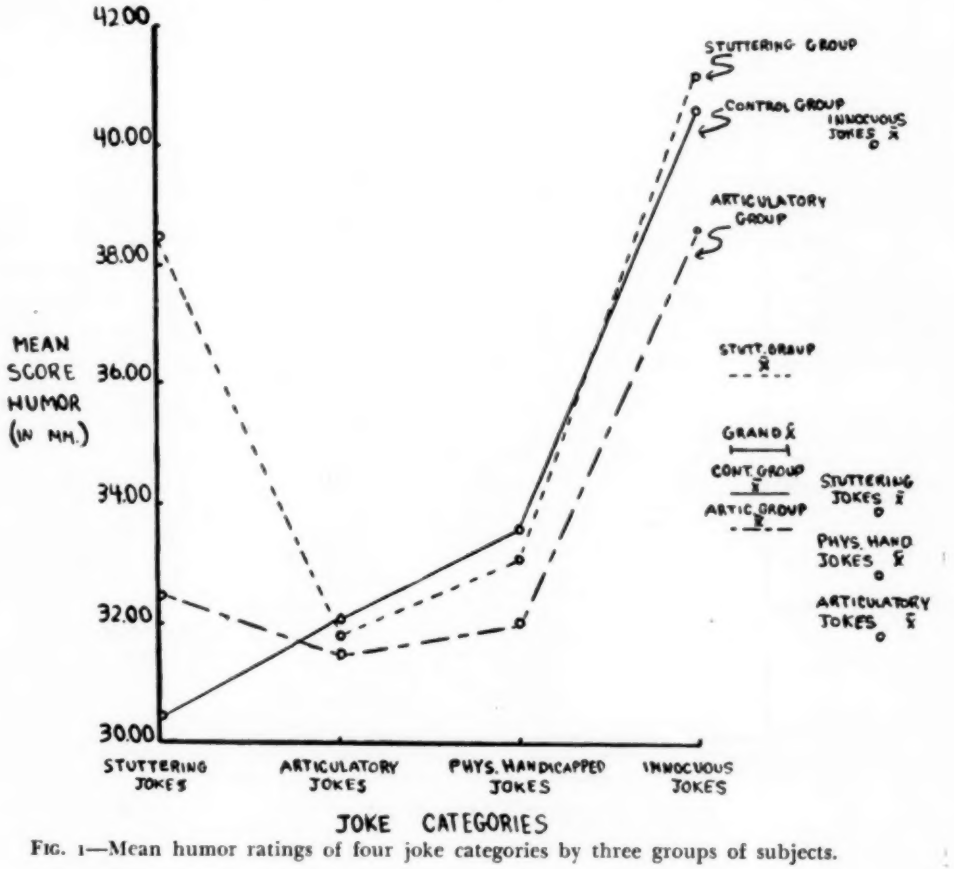

Fig. 1—Mean humor ratings of four joke categories by three groups of subjects.

2. The groups do differ significantly ($p = .02$) in their pattern of responses to the four categories of jokes. That is, the interaction of groups with joke categories is significant.

3. The joke categories differ significantly from each other. The subjects rated innocuous jokes as significantly more humorous than any category of jokes based on the handicapped.

4. Although the three groups are similar in their overall ratings of humor, there is a large difference among them on their ratings of jokes relating to stutterers. The stuttering group purports to find the greatest amount of humor and the control group the least amount of humor in such jokes. The interview data show the reverse of this. In the interview data, sixteen stutterers stated specifically that they disliked jokes about the handicapped, whereas only two of the control group stated that they disliked such jokes.

5. The three groups do not differ significantly from one another in their ratings of humiliation. However, the joke categories do differ significantly. That is, the degrees of humiliation found by the subjects in the four categories of jokes do differ significantly (See Figure 2 for results 5, 6, 7).

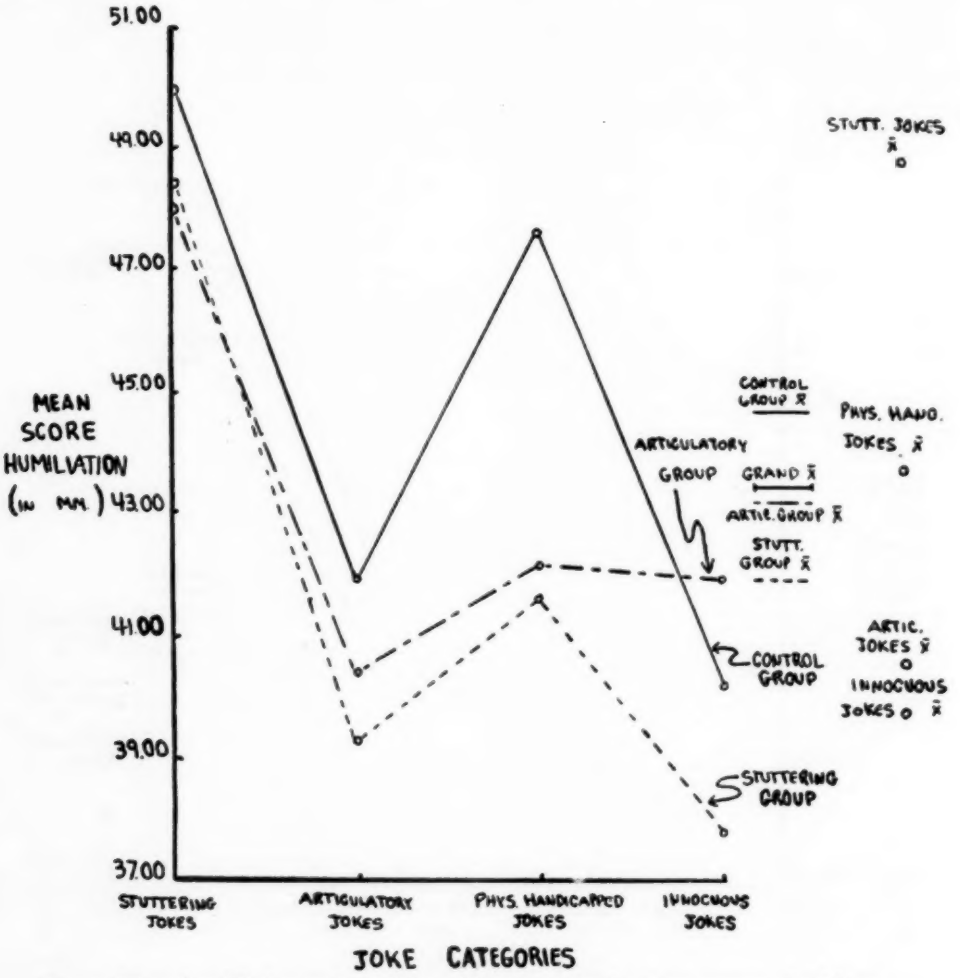

FIG. 2—Mean humiliation ratings of four joke categories by three groups of subjects.

6. The three groups do not rate the four categories of jokes significantly differently for humiliation. That is, the group profiles do not differ on the ratings for humiliation as they did on the ratings for humor. The lack of significance of the interaction of groups with categories can best be understood when seen graphically. The three profiles show considerable similarity.

7. Jokes directed toward stutterers are found to be most humiliating of all of the categories of jokes for all groups. This joke category differs significantly in humiliation from all other categories including the physically handicapped.

8. All sources of variance (groups, joke categories, and groups x categories) are significant ($p < .01$) for the data on "which joke I would not tell before the person who is the butt of the joke" (See Figure 3 for results 8, 9, 10, 11).

9. The control group differs significantly from both the articulatory and stuttering groups on whether they would tell certain jokes to a group in which a person was present who was like the character in the jokes. In all groups, however, jokes based on stuttering were rated as least likely to be told.

10. The stuttering and articulatory groups do not differ significantly from each other in regard to the question of whether they would tell the joke to a group in which a person like the principal character of the joke was present.

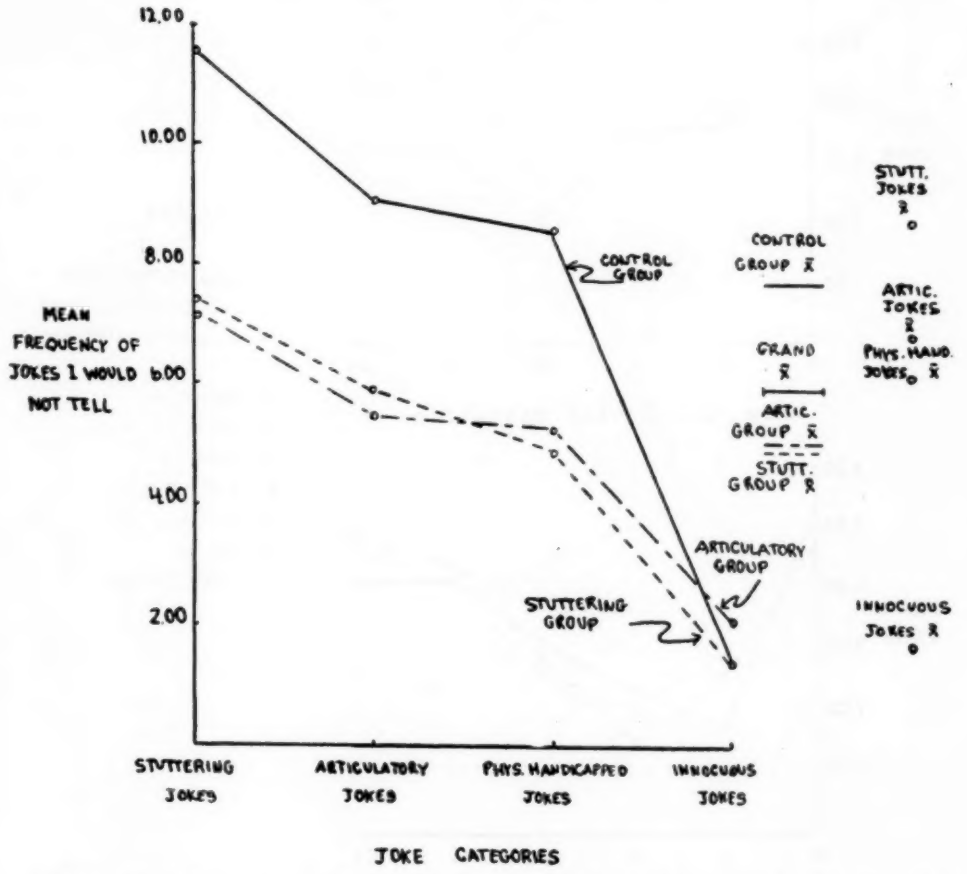

FIG. 3—Mean frequency of "Jokes I would not tell" by three groups of subjects on four joke categories.

11. For these data (which joke I would not tell before the person who is the butt of the joke) all pairs of joke category means differ significantly from one another ($p < .001$) with the single exception of the difference between the category of jokes based on articulatory defects and the category of jokes based on the physically handicapped.

12. In social situations stuttering is ranked as the most handicapping of the defects listed. Hearing and visual defects rank second and third. Orthopedic handicaps, along with lisping and sound substitutions, are regarded by all groups as being less handicapping (See Figure 4).

13. In school situations, hearing and visual defects rank first, with orthopedic disability once again being ranked as not so serious a handicap (See Figure 5).

14. In professional occupational situations, stuttering is again ranked as most handicapping with orthopedic disabilities ranked as least handicapping (See Figure 7).

15. Only in the laborer occupational situation does an orthopedic handicap become most serious. Even in this situation, stuttering ranks fourth in seriousness (See Figure 6).

III. Analysis of the Interview Data

Inconsistencies between the ratings on both humor and humiliation and the

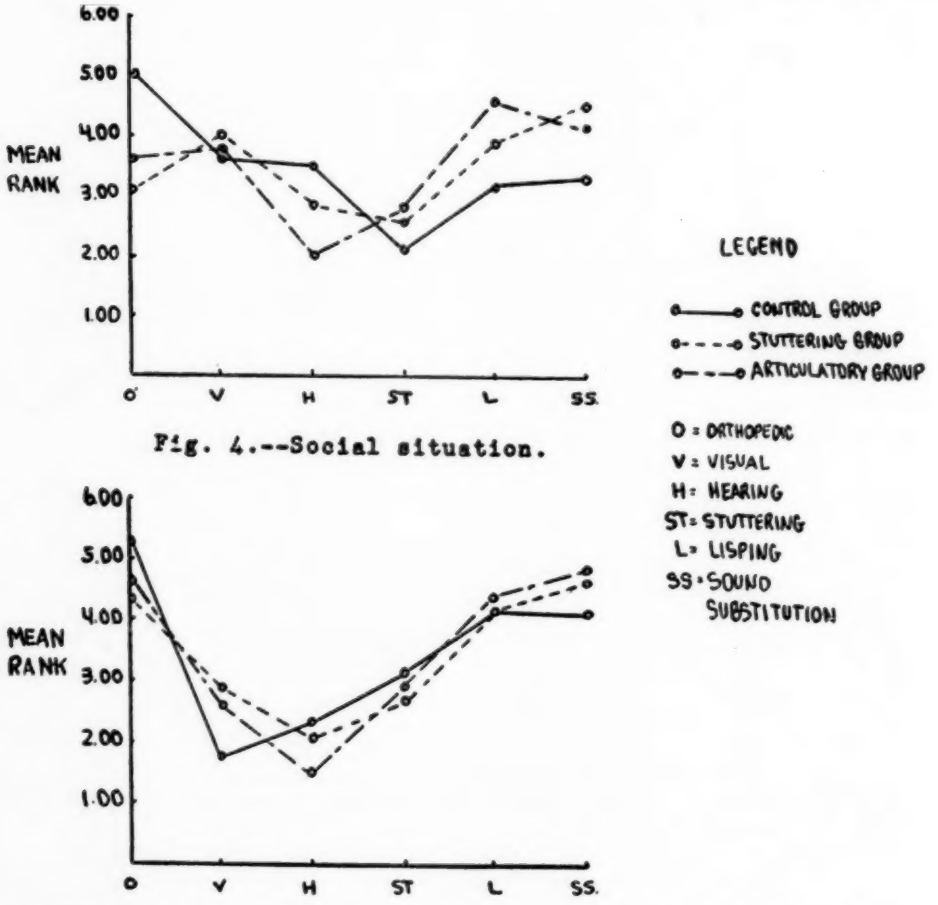

Fig. 4.--Social situation.

Fig. 5—School Situation. Mean Disability Rankings of Six Handicaps by Three Groups of Subjects in Two Situations—Social and School.

statements made in the interview are apparent. For example, stutterers rated jokes based on stuttering as being high in humor, but in the interview they said they disliked such jokes. Control subjects rated jokes on stuttering low in humor, but in the interview they indicated that they liked such jokes. Similarly, the ratings on humiliation by the subjects in the stuttering group and the subjects in the control group were disavowed in the interview. Other such facts brought out in the interview are summarized below.

There was no apparent relationship of the rating of the degree of humiliation to the person who was the butt of the joke, or the person underlined in the joke, with the rater's willingness to tell that joke in a group in which a person similar to the butt of the joke was present. A joke might not be told if it pointed up a defect even if it had little humiliation for the character within the joke, but a joke which was thought to be extremely humiliating might be told if it were also considered very funny.

According to the subjects in the study, when a joke is extremely funny, humiliation to the butt of the joke is lessened. Attention seems to be on the humor and not on the handicap.

Social pressures seem to act in two

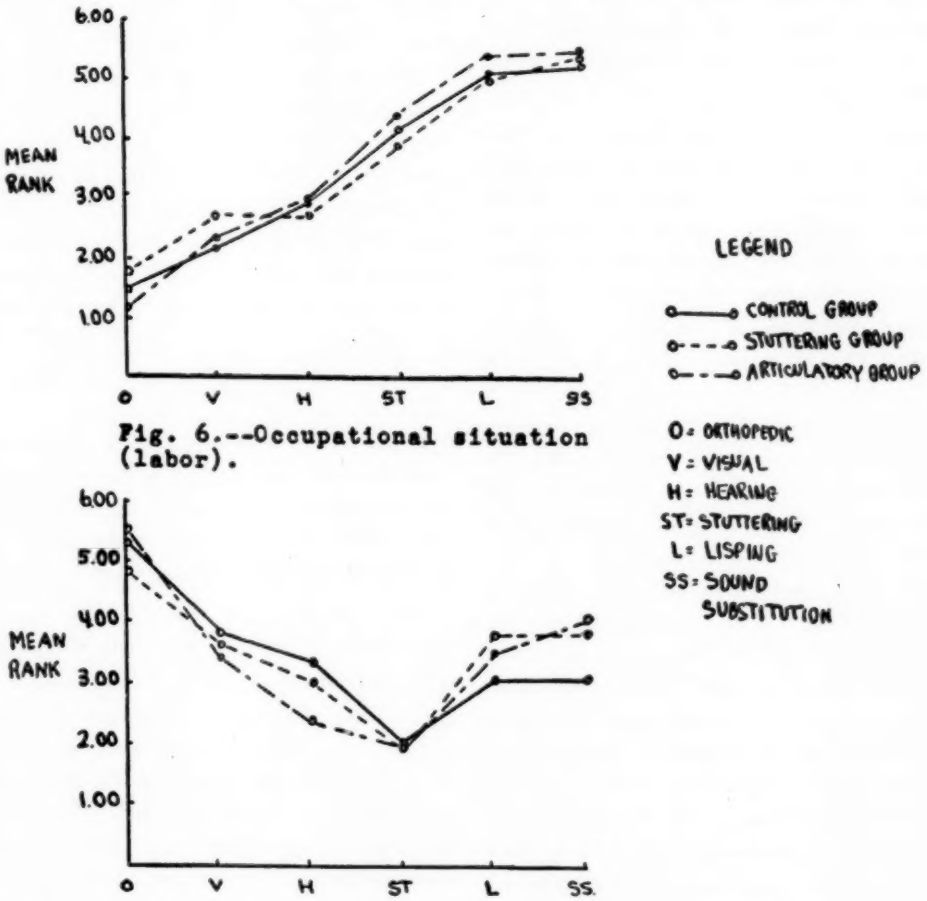

Fig. 6.--Occupational situation (labor).

Fig. 7—Occupational Situation (professional). Mean Disability Rankings of Six Handicaps by Three Groups of Subjects in two occupational situations—labor and professional.

ways: although the joke about a handicap may strike a person as funny, he feels somewhat guilty when laughing at it. And yet, he also feels that he should laugh at a joke concerning his own weaknesses because he feels this may indicate that a good adjustment has been made. This may also be due in part to the fact that our culture places a high premium on a good sense of humor.

It is apparent that among all of the members of these groups there are many conditions which limit the appreciation of a joke which features their particular weaknesses. Such appreciation appears to be influenced by: who is telling the joke, whether the joke is genuinely funny, the group in which it is told, and the intent behind the joke.

The stuttering group seemed to differ strongly from the other groups in indicating that they did not care for jokes about stuttering; at least they stated without being questioned that their degree of dislike for jokes based on stuttering certainly depended upon the joke, when it was told, who told it, the intent behind the joke, and whether the joke was genuinely funny.

Orthopedically handicapped people, as well as people with hearing defects, are considered by the subjects in the study to be "used to their defects" and not to mind their handicaps.

All three groups of subjects consider that orthopedic and hearing handicaps are not the fault of the person involved and are therefore less humiliating than handicaps such as speech defects, which are considered correctable. The feeling was frequently expressed or implied that a defect like stuttering includes psychological abnormality. No such psychological implications were attributed to orthopedic handicaps.

There were frequent statements that orthopedic and hearing handicaps occur only in older people and that these older people expect such handicaps and hence adjust more readily than younger people. No subject mentioned the possibility of speech defects in deaf people.

The control group seemed to understand the handicapped less fully than the stuttering and articulatory groups. It is doubtful, however, that the articulatory and stuttering groups really understand handicaps other than their own, with perhaps the stuttering group being more sensitive to all handicaps than the articulatory group. However, there appears to be little understanding on the part of the articulatory and stuttering groups of how the population at large feel about handicaps.

No subject could be found who was willing to doubt his ability to tell how sensitive another person would be to a joke. The subjects rated the jokes with confidence, and in the interview they tended to be dogmatic in their observations of how other people would respond to the humor and humiliation inherent in the jokes.

IV. Discussion

From the results of the material obtained from the rating scale and from the informal interview, as well as from the impressions of the investigator, it is possible to make some generalizations.

The Stuttering Group. In the beginning, let us state unequivocally that stutterers do not like jokes based on stuttering. This conclusion becomes inescapable in the light of the interview data even though the rating data indicated the contrary. Stutterers feel their disability strongly. They feel that they are not understood by others. They tend to be sensitive and easily hurt, but they often show a facade of hardness, and

even bravado, as a reaction to this sensitivity.

The stutterers feel that they are pitied but they respond to such pity with conflicting feelings. They sometimes want pity and accept it, for they are hurt. They often do not want it and resent it, for to accept sympathy and pity is to reveal sensitivity.

Stutterers feel that others sometimes laugh at them and they resent it. It is hard for a stutterer to realize that the one who laughs may feel inadequate and that the laughter represents a nervous release of tension rather than ridicule. Stutterers easily lose respect for those who tell jokes based on stuttering, or even for those who laugh at such jokes.

Stutterers feel that they receive more social disapproval than those who are physically handicapped. This is because stutterers sense an element of "psychological weakness" in their disability. They have failed to learn to speak correctly and they attribute their failure to inner weakness.

All of these attitudes and tendencies in stutterers combine to create group feeling. Stutterers tend to identify as a group, and in many respects we may accept the analogy to a minority group. There are strong bonds among the members of this group, and like members of other minority groups, they accept such status, for they find strength and solidarity in the group, although they reject the implications of this status.

The Articulatory Group. In contrast to the stutterers, the subjects with articulatory defects do not present homogeneous attitudes. Those in the articulatory group who feel that they have a disability as such (regardless of whether it is severe) tend to develop attitudes similar to stutterers. Others in this group feel that their defects are slight and are perhaps more akin to a blemish than a disability. They feel, too, with some justification, that the public in general does not regard such defects as constituting a disability.

The subjects in the articulatory group do not show commonality of attitude in their dislike for jokes based on articulatory defects. In the interview, stutterers were more consistent and clear in their attitudes; subjects with articulatory defects showed no such agreement. They presented a conflicting picture with different subjects holding different opinions and going different ways. They might be described as having taken a middle-of-the-road stand.

The subject with an articulatory defect lacks the identification with a group that was so apparent among stutterers. They tend, in general, to be more like the control subjects in their attitudes. Nevertheless, it is the impression of the investigator that the subjects in this group tended to hide their feelings. The evidence is nebulous, but the impression persists that these subjects, too, had unresolved tensions and anxieties, and felt unsure of their status in the general population. This may be because the articulatory defective subject more or less consciously recognizes that he has a defect which he does not disclose to others, feeling that others do not readily recognize his defect and that he may "get by." To continue our analogy to the minority group, then, persons with articulatory defects may be thought of as "passers." Frequently they can pass as normal individuals, but they always run the risk of being caught and labeled as having a disability.

The Control Group. The subjects in the control group, like those in the articulatory group, tended to have conflicting attitudes. For some subjects a handicap was lamentable and warranted sym-

pathy; other subjects tended to be hard and cold in their attitudes, although the attitude was often presented under the guise of being realistic.

The control subjects did not have a very clear understanding of the handicapped. They did not understand their problems and in some cases there was little interest in such problems. Many of the control subjects doubted that the attitudes of non-handicapped individuals toward the handicapped could be changed and some even questioned whether such changes were warranted.

If we may accept these attitudes as representative, the conclusion is clear that the non-handicapped individual tends to feel that it is incumbent upon the handicapped to adjust to society rather than the reverse. Indeed, many control subjects believed that the handicapped should become accustomed to hearing jokes based on their defects. However, subjects in the control group did not like jokes based on their own weaknesses.

In short, the control subjects tended to behave like the majority group which indeed they are. They tended to be partially compassionate and partially unfeeling and unsympathetic. They dislike jokes based on defects when such jokes are told in a group which includes a person with such a defect. This is partially due, perhaps, to guilt and partially due to a social consciousness which many of the control subjects possessed. But, in the interview, when they analyzed their feelings, hostility, insensitivity, and superiority came out.

V. Conclusions

The implications of these findings are many. There is, of course, much education to be done. The general population unquestionably has many misconceptions regarding the handicapped in general and the speech handicapped in particular. The subjects in this study, for example, frequently implied that orthopedic and hearing handicaps tended to be restricted to older people; that deafness could be corrected by hearing aids if you could get people to wear them; that articulatory defects are always correctable; that stuttering is curable. The non-handicapped person should understand, too, that although a handicapped person may laugh at jokes based on his weakness, he does not necessarily think they are humorous. People are frequently alienated through the use of such humor. In general, how such a joke is received will depend upon the situation as it is perceived by the handicapped person.

The handicapped are in need of education, too. They should be made aware of the true nature of their disability, as well as the attitude of the general population toward the disability. Clearly, it is unhealthy for the stutterer to regard his affliction as evidence of inner weakness and it is unrealistic and unwise for the person with an articulatory defect to pretend to others that no defect exists if he, himself, feels that a defect does exist.

Speech clinicians and teachers in general may find value in the results of this study, both in their relationship with the handicapped individual and in aiding the general population to a better understanding of the handicapped. Good rapport between student and teacher, or student and clinician, is essential to a profitable learning situation and sensitivity to the attitudes of the handicapped is basic to such a relationship.

Speech clinicians should realize too that they are not eradicating basic attitudes of the speech handicapped in

the hardening process frequently used to help the speech handicapped adjust to their disability. Indeed, it is a moot question whether such a hardening process in which basic attitudes are merely hidden is particularly healthy or leads to better adjustment.

And finally, for those media in which humor based on the handicapped is so prevalent and so widespread—the comics, the radio, television, and the moving pictures—much can be learned. If the results found in this study are representative, it is clear that innocuous jokes may produce more humor than jokes based on disabilities and, more important, would increase the appreciative audience to include the handicapped population.

A COMPARISON OF THE CHAIN ASSOCIATIONS OF NURSERY SCHOOL AND KINDERGARTEN CHILDREN TO ACTION-PICTURE STIMULI*

MARIAN NELSON FERGUSON
Arlington, Virginia

THE following study undertakes to investigate not only the comparative language development between the ages of three and six years of age but also some of the relations between thought and language in preschool children. Basing itself upon the association theory that processes related in the world are likewise related in the mind, the paper attempts to analyze these relationships by studying the language responses of children. Therefore, the associations of the children serve two purposes. First, they represent a language sample which can be analyzed and categorized to indicate the language development that takes place within the age range of the subjects. Secondly, as association responses they are indicative of underlying thought processes, and as such they can be studied from the standpoint of psycholinguistics.

I. Experimental Hypotheses

It is obvious that the language of the child increases in its complexity with increasing age and, through the use of two separate age groups of subjects, the present study attempts to trace the direction of this development within the age limits of 2:10 to 5:11 years. The hypothesis is *that within this age limit there is a development in the complexity of verbal behavior which can be measured by the frequency with which certain parts of speech are used.* In terms of linguistic science, the study is based on the classification of parts of speech rather than on a detailed linguistic analysis of the data, because the author was more familiar with the conventional grammatical categories.

No definite hypotheses can be made in the psycholinguistic area of investigation, but it was felt that certain tendencies might be noted concerning the relationship between language and thought on the association level. These were that children's associations tend to be indicative of their underlying thought processes, that selectivity of stimulus tends to produce a higher level of association than does perception of the total aspects of the stimulus picture, and that children's associations tend to involve more symbolic activity than do adult associations.

So far as the test instrument itself was concerned, an attempt was made to explore the use of "action-picture associations" in measuring language development. The hypothesis was *that action-pictures used as stimuli tend to produce a larger and more representative sampling of a child's verbal behavior than do still-pictures.* "Action-pictures" were defined as those pictures which depict objects and/or persons functioning in certain situations, a "still-picture" as

*Based upon M.A. thesis, Cornell University, 1954. The author wishes to acknowledge Professor R. W. Albright's help in completing the research and in the preparation of this report.

merely a picture of a single object without movement or action. In an action-picture, for instance, the picture of "door" is in a barnyard setting, in which a small boy is opening the barn door, whereas in a still-picture, there would be only a picture of a door unrelated to any action or setting. To measure the possible difference between "still" and "action" pictures, the present experiment was organized in a manner very similar to the Albrights' study in which still-pictures were used, so that a direct comparison could be made with their results.[1]

II. Experimental Methods and Procedures

Subjects: Of the 60 children tested, there were 30 in the nursery group, 15 boys and 15 girls, ranging in age from 34 months (2:10) to 60 months (5:0), the mean age being 49.8 months (4:2). The kindergarten group consisted of 30 children also, ranging in age from 61 months (5:1) to 71 months (5:11), the mean age being 64.1 months (5:4). The ratio of girls to boys (15:15 and 14:16 respectively) was not exactly equal, but it was felt that the ratio was close enough to be adequate for comparative purposes. Other studies have shown no significant sex differences between the free-associations of boys and girls in this age group.[2] It was presumed that the significant comparisons between the two age levels might point to the direction of development within this age range. The age criterion was used in this study only because of its convenience and because it is the criterion which the schools used to classify the subjects. Admittedly, age is an inadequate criterion for estimating linguistic or non-linguistic aspects of child development, since each child varies so much in maturational periods.[3]

There was little opportunity for making a critical selection of test subjects. Children were selected on the basis of age and accessibility, i.e., whether they were part of the nursery school and kindergarten groups tested. Intelligence quotients were not available for any of the children, but it is assumed that all of the subjects were of average or above average intelligence, since they were enrolled in the "normal" classes of the Ithaca public and private schools. According to their records, all the children were also free from any mental, physical, or speech handicaps, which might impair their ability to respond to the picture stimuli.

Procedure: The association test data was collected by means of recording the responses to a group of ten pictures, chosen in respect to two criteria: (1) each stimulus-picture of the Albrights' study was made the central stimulus of the more complex pictures used in this study, and (2) the picture included an action setting for the stimulus. The pictures of a *horse, bed, train, hand, table, house, book, spoon, door,* and *boy* were intended to be representative of some phase of the child's life. In all of the pictures used in the present study, the stimulus object of the original study made up a predominant part of the action-picture with but four exceptions: "spoon," "book," "door," and "hand" are more instruments than agents in the

[1] R. W. Albright and J. B. Albright, "Chain Associations of Preschool Children to Picture Stimuli," *Journal of Genetic Psychology*, LXXIX (1951), 77-93.

[2] Albright and Albright, (unpublished data).

[3] See the unpublished dissertation (Cornell, 1953) by D. Smillie, "Concept Formation and its Relation to the Social Behavior of Nursery School Children," p. 69.

actions depicted, and are therefore of secondary importance in the total pattern of the picture.

Each test was conducted individually in a private room within the familiar surroundings of the subject's own school, and another group of pictures was shown as a part of a "word game" by the regular teacher prior to the actual testing, so that the children would be familiar with the procedure which was to be used in the experimental situation. The author tested each child individually, in order to assure greater uniformity in giving the test instructions and in recording the responses, and to assure more constant social and personal relationships between the experimenter and each subject.

The directions given to each child were essentially the same as those used by the Albrights in their study. Specific instructions were elaborated on in some cases in order to make sure that each child understod how to play the "game." In this, the author was guided by the opinion of McNemar, who states that test understanding is more important than the standardization of practice intervals or other variables in establishing constant experimental conditions.[4] Strict time limits on associations were not imposed, although a limit of thirty seconds was adequate for nearly all of the chain responses to each picture stimulus.

III. Results

The responses obtained were subjected to two types of classification, quantitative and qualitative, the difference being in the amount of response for the former and the type of response for the latter. The quantitative classification, or the total number of responses given, is presented in Table I. It was found that the kindergarten group used a total of 4,055 words and the nursery 3,397, giving a sum total of 7,452 words elicited. The average length of response for each kindergarten subject on all ten pictures, then, was 135 words, whereas

TABLE I
Total Number of Response Words and Their Means in the Kindergarten and Nursery Groups

Picture Subject	Total Number of Response Words		Mean Number of Words	
	Kindergarten	Nursery	Kindergarten	Nursery
Horse	416	363	13.80	11.83
Bed	393	335	13.33	11.14
Train	379	357	12.63	11.79
Hand	367	257	12.23	8.55
Table	394	315	13.13	10.55
House	439	351	14.63	11.72
Book	440	381	14.67	12.66
Spoon	387	358	12.90	11.86
Door	429	356	14.30	11.79
Boy	411	324	13.70	10.86
Total Responses	4,055	3,397		
Average Number of Responses Per Child			13.53	11.28

[4] Q. McNemar, "Twin Resemblances in Motor Skills and the Effect of Practice Thereon," *Journal of Genetic Psychology*, XIIL (1933), 98-99.

the nursery school average was only 113 words, the combined average being 124 words. By comparison, the total response in the Albrights' study, which used still-pictures, was only 2,968 words, with an average total response of only 50 words per child. It would seem, therefore, that action-pictures elicit roughly twice as many responses as do still-pictures.

Before presenting the results of a qualitative analysis of the data, some discussion concerning the methods of classification is necessary. The words were classified according to the traditional grammatical categories rather than the form-classes of the linguists, as mentioned previously. There was little difficulty in judging the responses, since most of them were used within a phrasal context. One advantage of this action-picture test, in fact, is that it makes classification of grammatical forms comparatively easier, because the children tend to give more complex, phrasal responses, and a form can thus be classified more accurately in the context of a complete phrase or sentence.

TABLE II
PERCENTAGE OF WORDS IN EACH GRAMMATICAL CATEGORY OF THE TOTAL NUMBER OF RESPONSE WORDS IN EACH OF THE TWO STUDIES

	Present Study				Albrights' Study			
	Kindergarten		Nursery		Kindergarten		Nursery	
Grammatical Categories	Number of Words	Percentage of Total	Number of Words	Percentage of Total	Number of Words	Percentage of Total	Number of Words	Percentage of Total
Nouns	1,411	34.80	1,142	33.62	959	68.21	1,209	77.40
Verbs	338	8.33	269	7.92	138	9.82	121	7.75
Infinitives	41	1.01	26	.77	...	....	....	....
Other Verb Forms	189	4.66	185	5.45	158	11.24	54	3.46
Adjectives	212	5.23	125	3.68	29	2.06	31	1.98
Adverbs	124	3.06	96	2.83	11	.78	17	1.09
Compounds	94	2.32	60	1.77	61	4.34	45	2.88
Noun Genitives	11	.27	6	.18	2	.14	0	.00
Pronouns	371	9.15	320	9.42	0	.00	2	.13
Prepositions	302	7.45	206	6.06	...	....	....	....
Words Not Categorized*	962	23.72	962	28.30	48	3.41	83	5.31
Total	4,055	100.00	3,397	100.00	1,406	100.00	1,562	100.00

*Explanation of the words not categorized is made in the text.

TABLE III
SIGNIFICANT DIFFERENCES IN GRAMMATICAL USAGE BETWEEN THE KINDERGARTEN AND NURSERY SCHOOL GROUPS

Grammatical Categories	Kindergarten Mean	Nursery Mean	Critical Ratio of the Difference
Nouns	47.03	38.07	3.46**
Verbs	11.27	8.97	1.09
Infinitives	1.37	.87	1.14
Other Verbal Forms	6.30	6.17	.14
Adjectives	7.07	4.17	2.38*
Adverbs	4.13	3.20	1.04
Compounds	3.13	2.00	.24
Noun Genitives	.37	.20	1.21
Pronouns	12.37	10.67	.26
Prepositions	10.07	6.87	2.64*
Phrases	14.90	12.67	1.47

*Significant at the 5% level
**Significant at the 1% level

In all, 2,553 nouns were used (with the exception of those which were classified in the noun genitive and compound categories), 1,411 nouns in the kindergarten and 1,142 in the nursery. This was the largest single classification, including about 34 percent of the total responses for each group (see Tables II and III.) In addition, this was the only grammatical category that was significantly different at the one percent level between the kindergarten and nursery groups, meaning that only one percent of the time would the differences obtained be due to chance.[5] A total of 314 *different* noun responses were elicited, the kindergarten using 254 and the nursery 180. There were, however, only 118 "unique" responses in the combined groups, 72 of them being used in the kindergarten and 46 in the nursery school. "Unique" responses are those which were used only once in the combined responses of both groups, and the large preponderance of these in the kindergarten group might be indicative of the greater range of associations in the older group.

Of the 314 different noun responses used by both groups in the present study, only 118 were contained in Thorndike's list of the first 1,000 words.[6] This rather surprising difference between the present results and the Thorndike list may be indicative of several factors: (1) Thorndike's first 1,000 words take into account many of the "little" words, which would occupy a great proportion of the first 1,000 words; (2) the Thorndike list is based on samples of written language, which are less representative than samples of spoken language; and (3) the Thorndike list may be more conservative than is necessary in regard to children's vocabulary usage and understanding. Yet Thorndike's complete list on juvenile literature word frequency contained all but eleven of the noun responses given, which speaks well for its general adequacy.

An analysis of the total number of response words revealed that there was also a statistical significance at the one percent level between the total number of response words given by the two groups in favor of the older, kindergarten children. It is interesting to note here that the form most frequently used by both groups, i.e., nouns, was the one which showed the greatest difference between them.

The explanation of the highly significant difference in noun usage between the kindergarten and nursery would seem to lie in (1) the greater size of the kindergarten group's mean vocabulary, (2) the greater length of continuous, or phrasal, response, thereby increasing the occurrence of noun usage, and/or (3) the greater attention paid to details of the picture by the older group. In comparison, there was an even larger proportion of noun responses in the Albrights' study, accounting for almost 73 per cent of their total number of responses. This might indiate that still-picture associations are of a more descriptive nature, involving fewer continuous phrasal utterances than those elicited by action-picture stimuli. Yet, in both studies, a heavy preponderance of nouns might be expected, since nouns have been found to be the most frequent grammatical form used in any English language sample.[7]

[5] For a discussion of the formula for t used in this and other estimates of the significance of differences see W. J. Dixon and F. J. Nassey, Jr., *Introduction to Statistical Analysis* (New York, 1951).

[6] E. L. Thorndike and I. Lorge, *The Teacher's Word Book of 30,000 Words* (New York, 1944), 267-274.

[7] J. F. Markey, *The Symbolic Process and Its Integration in Children* (New York, 1928), 86-88.

Only those verbs which were the predicates of complete sentences were included in the "verbs" category, the other forms being placed in one of the other two verbal categories. The verbs were counted as forms, not as individual words, for this category: the response "has been eating" was counted as only one form. Such phrases were counted, however, as separate words in the quantitative analysis. The lower percentage of predicate verbs used by the nursery group, (Table II), although not statistically significant, would indicate a tendency of the younger group to use more incomplete phrases, since this group used a larger percentage of "other verbal forms," a category made up principally of verbal nouns and participles. By comparison, the subjects in the Albright study used almost the same percentage of "verbs" to other forms. In fact, the percentage of all verbal forms obtained in the present experiment was lower than would be expected in view of the type of picture-stimuli presented. The action-pictures, instead of eliciting a comparatively larger number of "action-forms," i.e., verbs, elicited a larger number of other grammatical forms. Another interpretation would be that action-pictures tend to elicit more complex and longer phrasal responses, but that longer phrases do not necessarily contain more verbs per phrase.

The higher percentage of verbal predicates, in relation to the "other verbal forms" of the present study, indicates a tendency toward complete sentences, which agrees with the results of both Otis[8] and Nice,[9] who found complete sentence responses to be more typical of speech at this level. There was a definite reversal of this tendency in the Albrights' study; they found that sentence responses occurred but rarely in chain responses to "still-picture" stimuli.[10]

In classifying adjectives, the articles *a, an,* and *the* were excluded for two reasons; (1) due to the speed at which most responses were elicited, there is some doubt that all such incidental words were recorded, and (2) these words do not have much value as attributes of the nouns they precede. They were, however, counted in the quantitative analysis, and they, along with auxiliary verbs, do form a large part of the "words not categorized" in Table II. The statistical significance of the number of adjectives would again be indicative of a possible development in language complexity in this age range.

The pronouns represented the greatest difference between this study and that of the Albrights, although this category was not statistically significant in either experiment. The present results would tend to support Markey's point, i.e., that pronouns in general hold second place in usage after nouns.[11]

The significance of the difference between the nursery and the kindergarten in using prepositions probably represents the best indication of language development in this age range, although few authors have written about this particular aspect of language. The noun and adjective differences could be a factor of the type of stimuli introduced in any given situation, but prepositions are relationship words which do not refer directly to visible objects and would therefore seem to be reliable indicators of language development. The

[8] M. Otis, "A Study of Association in Defectives," *Journal of Educational Psychology,* VI (1915), p. 278.

[9] M. M. Nice, "Length of Sentences as a Criterion of a Child's Progress in Speech," *Journal of Educational Psychology,* XVI (1925), p. 378.

[10] Albright and Albright, p. 88.

[11] Markey, p. 87.

TABLE IV
RATIO OF GRAMMATICAL USAGE AS EXPRESSED BY THE NUMBER OF FORMS OF THE NUMERATOR PER HUNDRED FORMS OF THE DEMONINATOR

Ratio	Kindergarten	Nursery
AVQ (Adjective/Verb Quotient)	63	42
PR-PQ (Preposition/Phrase Quotient)	67	54
VPQ (Verb/Phrase Quotient)	76	71
OVF-PQ (Other Verbal Forms/Phrase Quotient)	42	49

significant difference in prepositional usage between the two groups indicates the possible importance of these forms in language development within this age range.

Another rather useful indicator of language trends, which was devised by Boder, is the "adjective-verb" quotient.[12] The quotient is obtained by finding the number of adjectives per 100 verbs, but it can also be used to compare any two grammatical groupings. The results of such comparisons in the present study can be seen in Table IV. Boder has theorized that since verbs are more active forms than adjectives in the English language, a comparison of adjectives to verbs would reveal a trend toward either a more static or active type of language: that is, as the number of adjectives per 100 verbs increases, the more "static" is the person's use of language, since more verbs indicate more "active" language processes. If such an inference is valid, the results of the present experiment show that the nursery school group tends to associate, linguistically, in a more "active" way than does the kindergarten. This agrees with the view that young children tend to use language for its basic purpose of social action.

This same quotient might also be used to indicate a comparison of the relative complexity of language usage by calculating the number of other forms per 100 phrases. According to this formula, the kindergarten group would seem to exhibit a greater complexity of verbal behavior, because they used more prepositions and verbs per phrase than did the nursery group. The greater number of "other verbal forms" per phrase than verb predicates for the nursery school again indicates that the nursery tended to use a less complete type of utterance than the older group, if other verbal forms are more incomplete than predicate verbs, as the present study assumed.

IV. PSYCHOLINGUISTIC OBSERVATIONS ON THE EXPERIMENTAL DATA

No definite statements can be made, nor can any criteria, such as the levels of significance, be set up to evaluate the psycholinguistic findings that are to be presented, but the author felt that some discussion of this aspect of language would perhaps be useful to future investigators. It should be emphasized, however, that these observations have not been subjected to experimental methods which would assure their validity, but it was felt that observation of subjective behavior could properly be made without violating the spirit of psychological investigation.

The findings indicate that those children who used word forms, especially nouns, which referred to many aspects of the picture were more prone to give "descriptive" responses to the pictures, while those who concerned themselves with only a few central objects in the picture tended to give associations which

[12] D. R. Boder, "The Adjective-Verb Quotient: a Contribution to the Psychology of Language," *Psychological Record*, III (1940), 310-343.

went beyond the immediate picture stimuli and represented more coherent chains of associations. These two types of association might be equated with Rapaport's "close" and "distant" associations, "close" ones sticking relatively close to the stimulus pictures, and "distant" ones showing less apparent connection with the stimuli presented.[13] Some typical examples of these two types of association, in response to the picture with a spoon for a central object, were the following: (1) "That's a little boy eating, and he has a spoon and a bib on; there's some potatoes and some meat . . . ;" (2) "The little boy's eating and then he'll go up to the bathroom. By that time he'll be ready to go to bed, because it's his dinner he's eating." The first example would be considered a "close" association, because it tends to stay closer to the stimuli in the picture; the second is clearly a more "distant" association, because it moves away from the picture itself and toward related aspects of it. It would seem that the "distant" type of association, which was given by many of the subjects from both groups, represents a higher level of thought than does the "close" association. The greater amount of symbolic activity becomes more and more evident as the responses become more distant.

Another factor which seemed to determine the type of response given was the relative degree of emphasis, linguistically, which the subjects placed on the various aspects of the stimulus picture. Although no attempt was made to definitely categorize the types of association, it appeared that those children who mentioned only a few central aspects of the picture stimuli were more prone to give the "distant" associations, whereas those who mentioned many aspects of the picture gave more descriptive, or "close," associations. Therefore, it would seem that selectivity of stimulus, i.e., noticing a few central aspects, is a very important factor in thought processes, as expressed through associations.

Such a tendency toward "distant" association in children of both age groups is indicative of two possible conclusions:

(1) That verbal behavior does express the underlying patterns of other forms of behavior (as the second example above indicated).

(2) That "close" association tends to be an automatic or descriptive procedure which does not involve higher symbolic levels of thought to as great an extent as does "distant" association.

Therefore, adults, who tend to give "close" associations, as previous investigation has established,[14] would seem to be operating on a lower symbolic level in their word associations than many of these children who yielded more "distant" associations and who, consequently, are associating on a higher symbolic level. As was pointed out by Hunter, however, this lowering of symbolic activity on the part of adults could very possibly be the result of their greater automatization of speech, which results in the lessening of symbolic activity.[15]

V. Summary and Conclusions

This experiment was conducted in order to determine the possible differences in verbal behavior between kinder-

[13] D. Rapaport, M. Gill, and R. Shaffer, *Diagnostic Psychological Testing* (Chicago, 1946), II, pp. 24, 28.

[14] For a discussion of the associations of adults and children, see G. H. Kent and A. J. Rosanoff, "A Study of Association in Insanity," *American Journal of Insanity*, LXVII (1910), 317-390, and H. Woodrow and F. Lowell, "Children's Association Frequency Tables," *Psychological Monographs*, XXII, LI (1916), 1-110.

[15] W. S. Hunter, "The Symbolic Process," *Psychological Review*, LXI (1924), 478-497.

TABLE IV
RATIO OF GRAMMATICAL USAGE AS EXPRESSED BY THE NUMBER OF FORMS OF THE NUMERATOR PER HUNDRED FORMS OF THE DEMONINATOR

Ratio	Kindergarten	Nursery
AVQ (Adjective/Verb Quotient)	63	42
PR-PQ (Preposition/Phrase Quotient)	67	54
VPQ (Verb/Phrase Quotient)	76	71
OVF-PQ (Other Verbal Forms/Phrase Quotient)	42	49

significant difference in prepositional usage between the two groups indicates the possible importance of these forms in language development within this age range.

Another rather useful indicator of language trends, which was devised by Boder, is the "adjective-verb" quotient.[12] The quotient is obtained by finding the number of adjectives per 100 verbs, but it can also be used to compare any two grammatical groupings. The results of such comparisons in the present study can be seen in Table IV. Boder has theorized that since verbs are more active forms than adjectives in the English language, a comparison of adjectives to verbs would reveal a trend toward either a more static or active type of language: that is, as the number of adjectives per 100 verbs increases, the more "static" is the person's use of language, since more verbs indicate more "active" language processes. If such an inference is valid, the results of the present experiment show that the nursery school group tends to associate, linguistically, in a more "active" way than does the kindergarten. This agrees with the view that young children tend to use language for its basic purpose of social action.

This same quotient might also be used to indicate a comparison of the relative complexity of language usage by calculating the number of other forms per 100 phrases. According to this formula, the kindergarten group would seem to exhibit a greater complexity of verbal behavior, because they used more prepositions and verbs per phrase than did the nursery group. The greater number of "other verbal forms" per phrase than verb predicates for the nursery school again indicates that the nursery tended to use a less complete type of utterance than the older group, if other verbal forms are more incomplete than predicate verbs, as the present study assumed.

IV. PSYCHOLINGUISTIC OBSERVATIONS ON THE EXPERIMENTAL DATA

No definite statements can be made, nor can any criteria, such as the levels of significance, be set up to evaluate the psycholinguistic findings that are to be presented, but the author felt that some discussion of this aspect of language would perhaps be useful to future investigators. It should be emphasized, however, that these observations have not been subjected to experimental methods which would assure their validity, but it was felt that observation of subjective behavior could properly be made without violating the spirit of psychological investigation.

The findings indicate that those children who used word forms, especially nouns, which referred to many aspects of the picture were more prone to give "descriptive" responses to the pictures, while those who concerned themselves with only a few central objects in the picture tended to give associations which

[12] D. R. Boder, "The Adjective-Verb Quotient: a Contribution to the Psychology of Language," *Psychological Record*, III (1940), 310-343.

went beyond the immediate picture stimuli and represented more coherent chains of associations. These two types of association might be equated with Rapaport's "close" and "distant" associations, "close" ones sticking relatively close to the stimulus pictures, and "distant" ones showing less apparent connection with the stimuli presented.[13] Some typical examples of these two types of association, in response to the picture with a spoon for a central object, were the following: (1) "That's a little boy eating, and he has a spoon and a bib on; there's some potatoes and some meat . . . ;" (2) "The little boy's eating and then he'll go up to the bathroom. By that time he'll be ready to go to bed, because it's his dinner he's eating." The first example would be considered a "close" association, because it tends to stay closer to the stimuli in the picture; the second is clearly a more "distant" association, because it moves away from the picture itself and toward related aspects of it. It would seem that the "distant" type of association, which was given by many of the subjects from both groups, represents a higher level of thought than does the "close" association. The greater amount of symbolic activity becomes more and more evident as the responses become more distant.

Another factor which seemed to determine the type of response given was the relative degree of emphasis, linguistically, which the subjects placed on the various aspects of the stimulus picture. Although no attempt was made to definitely categorize the types of association, it appeared that those children who mentioned only a few central aspects of the picture stimuli were more prone to give the "distant" associations, whereas those who mentioned many aspects of the picture gave more descriptive, or "close," associations. Therefore, it would seem that selectivity of stimulus, i.e., noticing a few central aspects, is a very important factor in thought processes, as expressed through associations.

Such a tendency toward "distant" association in children of both age groups is indicative of two possible conclusions:

(1) That verbal behavior does express the underlying patterns of other forms of behavior (as the second example above indicated).

(2) That "close" association tends to be an automatic or descriptive procedure which does not involve higher symbolic levels of thought to as great an extent as does "distant" association.

Therefore, adults, who tend to give "close" associations, as previous investigation has established,[14] would seem to be operating on a lower symbolic level in their word associations than many of these children who yielded more "distant" associations and who, consequently, are associating on a higher symbolic level. As was pointed out by Hunter, however, this lowering of symbolic activity on the part of adults could very possibly be the result of their greater automatization of speech, which results in the lessening of symbolic activity.[15]

V. Summary and Conclusions

This experiment was conducted in order to determine the possible differences in verbal behavior between kinder-

[13] D. Rapaport, M. Gill, and R. Shaffer, *Diagnostic Psychological Testing* (Chicago, 1946), II, pp. 24, 28.

[14] For a discussion of the associations of adults and children, see G. H. Kent and A. J. Rosanoff, "A Study of Association in Insanity," *American Journal of Insanity*, LXVII (1910), 317-390, and H. Woodrow and F. Lowell, "Children's Association Frequency Tables," *Psychological Monographs*, XXII, LI (1916), 1-110.

[15] W. S. Hunter, "The Symbolic Process," *Psychological Review*, LXI (1924), 478-497.

garten and preschool children. There was also an attempt to reveal more clearly the relationship between language and thought, as brought out by the associations of the children to action-picture stimuli.

A group of sixty subjects was given an association test, using action-pictures as stimuli. The age range of the subjects was from three to six years, and the language samples obtained were compared between the thirty kindergarten and thirty nursery school subjects in regard to the number and kinds of speech forms used, in order to trace the developmental aspects of their verbal behavior. It was found that the kindergarten group used a significantly larger number of nouns, adjectives, and prepositions, as well as a large total number of words. When the number of various speech forms were compared with the number of phrases, the kindergarten group was found to be superior to the younger group, in terms of the complexity of phrasal structure, since it used more prepositions, nouns, and verbs per phrase than did the nursery group.

The results of this study were also compared with a similar study which used still-pictures as stimuli, so that the relative effectiveness of both types could be evaluated. Although a greater number and variety of speech forms were elicited by the children in response to the action-picture stimuli, the more active forms, i.e., verbal forms, did not show any significant increase in occurrence.

In conclusion, it can be said that within the age limit of approximately three to six years there is a development in the complexity of verbal behavior which may be measured by the frequency with which certain parts of speech were used. These significant differences between the two groups are as follows:

(1) That there is a significant difference at the one percent level between the younger and older children in both number of total responses and number of noun constructions.
(2) That there is a significant difference at the five percent level between them in respect to their use of adjectives and prepositions.

No definite conclusions can be reached in the psycholinguistic area of investigation, but certain tendencies can be noted concerning the relationships between language and thought on the association level. These tendencies are the following:

(1) That children's associations tend to be indicative of their underlying thought processes.
(2) That selectivity of stimulus tends to produce a higher level of association than does perception of the total aspects of the stimulus picture.
(3) That children's associations tend to involve more symbolic activity than do adult associations.

Finally, the results of the experiment indicate that action-pictures tend to produce a larger, and probably more representative, sampling of a child's verbal behavior than do still-pictures.

THE CONSTRUCTION AND TESTING OF A FORCED CHOICE SCALE FOR MEASURING SPEAKING ACHIEVEMENT*

KEITH BROOKS
Ohio State University

THE problem of assessment of speaking achievement is difficult and complex. The degrees of emphasis on any of the fundamental processes often vary from teacher to teacher and school to school. While the problem of assessment is difficult and complex, it certainly is not impossible. A great deal has been accomplished. Speech Rating Scales, relatively free from the usual rating scale weaknesses, have been developed and are in use. Psychological and Educational testing research has contributed greatly to the advance of rating scale construction and use.

I. The Problem

This study is concerned with the construction and testing of a Forced Choice Scale for measuring achievement in speaking. The theory behind the construction of a Forced Choice Scale is ascribed to Horst by Wherry.[1] The first published suggestion for the Forced Choice technique is quoted from Horst below:

> One of the greatest weaknesses of current temperament and personality tests is the fact that they can be "beaten" by persons with sufficient intelligence . . . However, with sufficient ingenuity and resourcefulness, it should be possible to develop techniques which would reduce considerably the possibilities of "beating" the questionnaire type of test. One method which is being experimented with by the National Research Council Committee on Selection and Training of Aircraft Pilots is to have items arranged in pairs in such a way that both members of the pair are about equal in general social acceptability and unacceptibility. However, each item is presumably on a different continuum and in order for the subject to give the correct choice, he must first identify the continuum. It is quite possible that this principle, or modifications of it, can be extended and further developed. The general principle here is to require the subject to express a preference for one of two alternatives when each member of a pair is equally derogatory or complimentary. It is believed that if systematic effort in this field were encouraged, it would be possible to develop other techniques which would enable much improvement in the validity of questionnaire responses.[2]

Wherry is accredited with the initial practical application of Horst's theory.[3] His work was reported in an abstract of a paper given at an American Psychological Association meeting by the "staff of the Personnel Research Section of the Adjutant General's Office."[4] Wherry's definition of the Forced Choice technique, as given in this paper, involves forcing

> . . . the rater to choose between descriptive phrases which appear of equal value for selection by the rater, but which are different in validity or discrimination value for the cri-

*Based upon Ph.D. dissertation, Ohio State University, 1955, directed by Franklin H. Knower.

[1] Leonary V. Gordon, "A Comparison of the Validities of the Forced Choice and Questionnaire Methods in Personality Measurement," Ph.D. dissertation, Ohio State University, 1950, p. 27.

[2] Paul Horst, *The Prediction of Personal Adjustment,* Bulletin Number 48, Social Science Research Council, 1941.

[3] Walter S. Monroe, *Encyclopedia of Educational Research,* New York, 1950.

[4] Staff of the Personnel Research Section of the Adjutant General's Office. "The Forced Choice Technique and Rating Scales," *The American Psychologist,* I (1946), p. 267.

terion—for example, military leadership. The major problem is the grouping of alternatives to achieve these ends.

Sisson's report on his work with the Forced Choice Scale indicated that its chief value was its ability to reduce conscious bias and to minimize individual differences in interpretation of qualitative adjectives and meaning of different scale points. Also, he found that this technique produced a better distribution of ratings in contrast to the negatively skewed leptokurtic distribution of ratings of older methods.[5]

The Staff of the Personnel Research Section of the AGO in reporting on the Forced Choice technique and Rating Scales stated that the utility of rating scales is limited by the ease with which the rater can determine where he is placing a person on the scale. The Forced Choice technique, they claim, is designed to reduce the rater's ability to control the final result of the rating.[6] The technique used in accomplishing this end involves forcing the rater to choose between descriptive phrases which appear of equal value—equal preference indices—but are different in validity—discrimination indices. A *preference index* is the mean of the scale values indicating the degree to which the criterion label applies to the group concerned. A *discrimination index* represents the correlation of the criterion label with an over-all rating.

The pairing of alternatives to achieve these ends involves five major steps: (1) procurement of descriptive phrases, (2) preparation of a complete list of such phrases, (3) determination of preference and discrimination indices, (4) pairing of alternatives as described above, and (5) application to a specified criterion group.

The problem of testing in this study involves a comparison between the Knower General Speech Performance Scale[7] (hereafter referred to as a Simple Numerical Scale) and the Forced Choice Scale developed in this study. So that the comparison of the two instruments might be principally methodological, it appeared advisable to construct the Forced Choice Scale from the same categories and, as far as possible, from the same items as already existed in the Knower form.

It is hoped that the development of a Forced Choice Scale for measuring achievement and for comparative purposes will throw some light on the differential operation of the two methods (Numerical Scale and Forced Choice Scale) and will lead to the further refinement of the Forced Choice Scale and later scales.

II. Procedure and Analysis of Scale Construction

In order to insure a scale with a vocabulary meaningful to the student population in the beginning speech fundamentals course, sixty speech students were asked to write descriptions of the observable speaking habits of effective and ineffective speakers in their speech classes. Thirty students described the observable speaking habits of any classmate considered to be one of the best speakers in their class and thirty students described the observable speaking habits of any classmate considered to be one of the poorest speakers in their class. The sixty descriptions concerned eighteen different students classified as among the

[5] E. D. Sisson, "Forced Choice, The New Army Rating," *Personnel Psychology*, I (1948), 365-381.

[6] Staff, Personnel Research Section, AGO, *op. cit.*

[7] Items in Knower General Speech Performance Scale were reported in *Psychological Studies of Individual Differences among Students of Speech* by H. Gilkinson and F. Knower, University of Minnesota, June, 1939, pp. 101-185.

best in the class and twenty-two different students classified as among the poorest in the class. Two hundred and forty-four descriptive phrases were culled from these essays. Leading speech fundamentals texts and available Speech Performance Check Lists were consulted for further descriptive phrases. A total of two hundred and sixty-one descriptive phrases of speaking habits (both desirable and undesirable) were collected. These items were next categorized under the seven general categories of emphasis in speaking as established in the Knower General Performance Speech Scale: Speech Attitudes and Adjustments, Voice and Articulation, Bodily Posture and Action, Language, Content—Speech Materials, Organization, and Audience Interests and Adaptation.

It was assumed at the start that no one of these categories was more important than another. Although certain of these categories often receive more emphasis than others, no conclusive evidence is available to indicate that such emphasis is desirable. It was, therefore, considered desirable to include as many items under each category as were considered applicable to that particular division.

The list of 261 items was next checked for ambiguity and repetition, and 133 items remained. Of these 133 items, 17 were listed under the heading of Attitudes and Adjustments, 26 were listed under the heading of Voice and Articulation, 28 were listed under the heading of Bodily Action, 12 were listed under the heading of Language, 21 were listed under the heading of Content—Speech Materials, 18 were listed under the heading of Organization, 11 were listed under the heading of Audience Interests and Adjustments.

Mimeographed check lists of the 133 items were prepared to be administered to 300 students in a beginning speech course. Each item was checked on a 1 to 5 point scale (slight degree, limited degree, typical degree, outstanding degree, exceedingly high degree) indicating the applicability of each item to the speaker the rater had in mind. Responses were recorded on IBM answer sheets. The following directions were mimeographed at the head of each check list of items.

Directions:

1) Write your name and your instructor's name on left corner of the answer sheet.
2) Identify in your mind those individuals in this class whom you consider to be the *best* speakers, the individuals in this class whom you consider to be the most *average* speakers, and the individuals in this class whom you consider to be the *poorest* speakers.
3) Select *one* speaker from these three groups whose speaking habits you feel best able to describe. *Write his name in the upper right hand corner of the answer sheet and indicate what group you feel he is in—best, average, poorest.*
4) Keeping this one individual in mind, describe him by indicating your responses to each item on the answer sheet. Your answer sheet has five spaces after each number. If you mark space 1, you have indicated that the item is applicable to the speaker you have in mind to a *slight degree*. Marking space 2 would indicate that the applicability of this item to the speaker you have in mind is to a *limited degree*. Space 3 indicates a *typical degree*. Space 4 indicates an *outstanding degree*. Space 5 indicates an *exceedingly high degree* of applicability of this item to the speaker you have in mind.
5) Remember that the speaker you are describing may be better on some items and poorer on others. Therefore, be sure that you consider each item individually as it applies to the speaker you have in mind.
6) Be sure that you are as accurate as possible.
7) You will be given 25 minutes.

The check list was submitted immediately after the mid-term examination, at which time it was felt that the students were sufficiently familiar with speech principles and student speakers in their classes to be able to critically describe the observable speaking habits of at least one member of the

class. The 100 students checking the list of 133 items in terms of a classmate they regarded as one of the best speakers in the class described 41 different student speakers. The 100 students checking the list of 133 items in terms of a classmate they regarded as one of the most average speakers in the class described 75 different student speakers. The 100 students checking the list of 133 items in terms of a classmate they regarded as one of the poorest speakers in the class described 56 different student speakers. In all, 300 students described 172 different classmates.

Each of the three categories of speakers rated on IBM answer sheets—100 student descriptions of best speakers, 100 student descriptions of average speakers, 100 student descriptions of poorest speakers—was run separately through the Graphic Item Counter on the International Test Scoring Machine. The board was wired so that preference counts for each of the five choices on the key for each item would be obtained. This permitted a count of 18 items for each wiring and necessitated eight wirings to obtain position counts for all 133 items in each of the three categories.

These preference counts were next used to determine a preference index and a discrimination index for each of the items. A sample of the statistical procedure followed is given in Table I.

This procedure was followed for each of the 133 items. To facilitate the pairing of items according to like preference indices and unlike discrimination indices the items were divided into two categories—effective speaking habits and ineffective speaking habits—and listed in descending order according to preference indices.

The use of the Preference and Discrimination Indices as aids in pairing the items posed two problems: (1) within what limits were varying preference indices to be regarded, for all practical purposes, as being alike, and (2) at what point did a significant difference between discrimination indices exist. The first problem was sufficiently answered by determinining the Standard Error of the Mean of each preference index. These Standard Errors indicated the possible range within which the mean (preference index) might occur two-thirds of the time in any like population. The second problem was solved by transforming all discrimination indices into validity r's (Pearson product moment correlation) and determining the fiducial limits of each r at the 5% and 1% levels.

On the basis of this informaiton *each*

TABLE I
STATISTICAL PROCEDURE FOR DETERMINING PREFERENCE AND DISCRIMINATION INDICES

Item 2: Confident in self					
Alternatives:	1	2	3	4	5
Weight (w):	0	1	2	3	4
Frequence (f):					
Best	0	0	8	41	51 = 100
Average	6	24	42	29	5 = 100
Poorest	32	49	15	2	2 = 100
Σ f	32	73	65	72	58 = 300
Σ w·f	0	73	130	216	232 = 651
d [B—P]	32	49	7	39	49 = 176

Preference Index: $\frac{f \cdot w}{n} \times 100 = \frac{651 \cdot 100}{300} = 217$

Discrimination Index: d = *176*

item was listed with all possible pairing choices. The possible choices included those items whose preference indices fell within the limits of the Standard Error of the preference index concerned and whose discrimination indices were significantly different at the 1% level from the discrimination index concerned. Many of the items listed under the heading of Effective Speaking Habits were also listed—in negative form—under the heading of Ineffective Speaking Habits. The selection of any one of these items for pairing eliminated its counterpart from being paired. This elimination prevented content duplication of items.

It was assumed preferable to pair items previously listed under the *same* category whenever possible. Due to varying degrees of emphasis on the seven categories it was felt that combining items from different categories *might* influence rater choice. If, for example, an individual felt that the category *Organization* was more important than the category *Voice and Articulation*, then he might exercise personal bias in item selection. By pairing items from the same category, this possibility might be avoided.

Another problem involved in the pairing concerned the doubtful advisability of combining extreme differences in discrimination indices. The possibility of obvious face value differences appearing between items with extreme differences in discrimination indices was considered. Although no statistical evidence was available to support this premise, it was felt that such a possibility should be avoided in pairing.

These two criteria—(1) pairing items from the same category, and (2) avoiding combinations of items with extreme differences in discrimination indices—were followed whenever possible. Fifty-two items were paired. Criterion 1 was possible in 10 of the 26 pairs. Criterion 2 was possible in 21 of the 26 pairs. In all cases, the pairings met the original statistical criterion of like preference indices and significantly different discrimination indices at the 1% level of confidence.

Tetrads were formed by combining pairs of effective speaking habits with pairs of ineffective speaking habits. This process completed the formation of the scale.

III. Procedures and Analyses of Scale Comparisons

The construction of the scale having been completed, a method of comparison with other existing scales was undertaken. A simple numerical scale was selected because it was the type commonly used in evaluation of achievement in speech courses. The writer was particularly interested in determining if a Forced Choice Scale was comparable to the simple Numerical Scale in terms of (1) usability and (2) efficiency. The following procedure was pursued to gather data on which to answer these questions.

The Forced Choice tetrads were arranged on a single sheet of paper with two boxes after each item allowing space for rater's checks. The scale directions were:

Directions: Check *two* items in each of the thirteen groups below, as follows:

(1) In the column headed MOST, check the *one* item in each group which is *most* applicable to the speaker.

(2) In the column headed LEAST, check the *one* item in each group which is *least* applicable to the speaker.

The responses were scored +1, 0 and −1 according to the discrimination value of each item in a tetrad. If an effective habit was selected as *most* applicable to the speaker it was scored +1 when it discriminated more (according

to discrimination indices) than the other item in the pair, *or* o when it discriminated less. If an effective habit was selected as *least* applicable to the speaker it was scored —1 when it discriminated more than the other item in the pair, *or* o when it discriminated less. If an ineffective habit was selected as *least* applicable to the speaker it was scored +1 when it discriminated more than the other item in the pair, *or* o when it discriminated less. If an ineffective habit was selected *most* applicable to a speaker it was scored —1 when it discriminated more than the other item in the pair, *or* o when it discriminated less. An example of a possible tetrad and its scoring is shown in Table II. The discrimination indices are listed to clarify the scoring. Note that the effective speaking habits have like preference indices and significantly different discrimination indices. The same is true of the second pair—the ineffective speaking habits.

Twenty student speakers gave five-minute Belief Speeches over a period of four days. Each day five speakers were rated by a like number of students using the Numerical Scales and a like number of students using the Forced Choice Scale. Students were allowed to check their scales while the speakers spoke; however, only two minutes were allowed to complete checking after each speaker concluded. All scales were collected two minutes after the conclusion of the last speech each day.

The total scores of all judges for each speaker were totaled to determine the rank order of speakers for each day. The charts for the first day are given in Table III.

The rank order of all speakers each day for four days as determined by the Forced Choice and Numerical Scales is

TABLE II
A POSSIBLE TETRAD AND ITS SCORING

	Most	Least	Discrimination	Preference
Speaks fluently	+1	—1	174	212
Good sentence structure	0	0	130	210
Excessive movement	0	0	70	77
Lacks information	—1	+1	136	73

TABLE III
RANK ORDER OF ALL SPEAKERS DETERMINED BY FORCED CHOICE SCALE AND SIMPLE NUMERICAL SCALE ON FIRST DAY

Forced Choice Scale

Speakers	Judges														Total	Rank
A	17	19	13	18	17	12	19	18	12	14	13	17	18	15	234	1
B	14	14	11	9	6	8	15	13	15	14	14	9	15	11	183	3
C	7	10	5	—2	5	5	6	12	9	3	12	1	11	11	81	5
D	12	5	9	12	10	11	13	11	8	5	—2	9	5	8	109	4
E	12	19	8	14	6	10	14	17	15	13	14	11	16	14	192	2

Simple Numerical Scale

A	36	27	27	32	32	30	31	34	30	34	30	33	26	35	468	1
B	37	31	23	29	28	28	28	32	31	35	31	31	27	37	457	3
C	31	24	19	26	24	23	25	25	25	30	26	25	22	30	379	5
D	34	23	21	27	25	27	27	31	27	32	23	28	23	32	406	4
E	35	30	25	30	30	31	30	35	29	33	28	36	28	36	466	2

TABLE IV
COMPARISON OF DAILY RANK ORDER OF SPEAKERS DETERMINED BY FORCED CHOICE AND NUMERICAL SCALE

Speakers	First Day		Second Day		Third Day		Fourth Day	
	FCS	NS	FCS	NS	FCS	NS	FCS	NS
A	1	1	1	1	3	3	4	4
B	3	3	5	5	5	5	3	3
C	5	5	2	2	1	1	1	2
D	4	4	3	3	4	4	5	5
E	2	2	4	4	2	2	2	1

TABLE V
INTERCORRELATION OF SCORES AND FIDUCIAL LIMITS

Trials	Judging Group	Forced Choice	Fiducial Limits—5%	Judging Group	Numerical Scale	Fiducial Limits—5%
1	B	.61	.51-.69	A	.80	.74-.85
2	A	.36	.20-.50	B	.43	.28-.56
3	B	.72	.63-.79	A	.90	.86-.93
4	A	.57	.47-.69	B	.68	.56-.75

shown in Table IV. On the first day 14 student judges used the Forced Choice Scale and 14 other students used the Numerical Scale. These 28 judges were concerned with the same five speakers. Five different speakers were used each day and the judges alternated daily in their use of the Forced Choice and Numerical Scale—both scales being used each day but by alternate groups.

The average rank order intercorrelations of scores of the judges using each scale each day are shown in Table V. The formula used was that for determining an average intercorrelation.[8] The fiducial limits were established according to Edwards, p. 136, with an adaptation of his standard errors formula as follows: $z^1 = \sqrt{(N-3)\ (n \cdot \frac{n-1}{2})}$. The fiducial limits were determined at the 5% level of confidence and showed a significant difference between the average r's two out of four times. Both of these times Group B was using the Forced Choice Scale and Group A was using the Numerical Scale. The average r increased each time Group A and Group B used the Forced Choice Scale and each time they used the Numerical Scale.

To deal with the eight average r's, they were changed to z' scores for normal distribution. The z' transformation for each of the average r's of the scores from the four uses of the Numerical Scale and from the four uses of the Forced Choice Scale were submitted to a t-test to determine the significance of the scores from the two methods. The resulting t-test of 2.835 was not statistically significant (d.f. 3; 10% level of confidence).

IV. DISCUSSION AND SUMMARY

On the basis of the procedures followed in this study in constructing a Forced Choice Scale for measuring speaking achievement and in comparing this scale to a Simple Numerical Scale, several observations appear to be valid.

The use of student-written descriptive essays on the observable speaking habits of best and poor student speakers produced a representative list of effective and ineffective observable speaking habits. The final list of 133 items provided the criteria by which 300 students in a beginning speech course each eval-

[8] Allen L. Edwards, *Experimental Design in Psychological Research* (New York, 1953), p. 409.

uated a particular classmate whom he felt was either one of the best, one of the most average or one of the poorest speakers in the class.

The technique used as a basis for calculating discrimination and preference indices was a satisfactory and convenient method. The resulting indices showed that 19 of the 133 items could not be paired due to preference indices that were significantly different from any other preference index and/or because of discrimination indices which were *not* significantly different from another discrimination index whose preference index was considered the same as the preference index of the item concerned. Of the remaining 114 items, 54 items were effective speaking habits and 60 items were ineffective speaking habits. Thirty-nine of the 54 effective habits were the counterparts of 39 of the 60 ineffective speaking habits. Consequently, if one of these effective habits was selected for pairing, its counterpart was automatically eliminated and vice-versa. Seventy-five items remained.

Twenty-three of the remaining 75 items were rejected for one or more of the following reasons:

(a) Obvious face-value difference in pairing possibilities.
(b) Extreme difference in discrimination indices.
(c) Expert opinion judged item to be of doubtful value or relevance to achievement.

Fifty-two items remained for pairing. The final form included 26 pairs—13 pairs of effective speaking habits and 13 pairs of ineffective speaking habits. Thirteen groups, each including two pair of phrases describing effective and ineffective speaking habits, were formed on a single sheet of paper.

A training period in the use of the Forced Choice Scale in the classroom is essential. One day was used for this purpose in this study. A second day is recommended and justified in terms of student training in critical evaluation.

The statistical procedure followed in comparing the Forced Choice Scale and the Simple Numerical Scale revealed the following:

(a) The rank orders of all speakers as determined by the Forced Choice Scale and by the Simple Numerical Scale were comparable.
(b) The average r increased each time either scale was used. The average r for the Numerical Scale was in every case higher than the average r for the Forced Choice Scale although only twice was it significantly different. An overall comparison of the two did not yield a significant t-ratio.

The statistics from the comparisons between the two scales must be regarded as only possible indications of the potential of the Forced Choice Scale, due to the following differences within the two scales compared:

(a) The Numerical Scale was familiar to the students, having been used by them for approximately two months. The Forced Choice Scale was given a one-day trial run.
(b) The Numerical Scale allows for manipulation of scores in determining final rank order of all speakers. Student bias can, therefore, determine ranks according to previous information about the speaker as revealed in class discussions, teacher criticisms, and grades. The Forced Choice Scale appears to give the rater no indication of scoring values.
(c) High intercorrelations of scores of the group of judges is possible by use of a Numerical Scale. Such correlations may be attributable to

either excellence of judgment or a consistent bias.

In conclusion:

1. In terms of efficiency, it was found that the Forced Choice Scale required no more time for student checking than did the Numerical Scale. Also, the Forced Choice Scale can be scored in approximately ten seconds. The Numerical Scale is self-scoring.

2. The Forced Choice Scale developed in this study has particular value as a research instrument where values of criteria need not be explained to students who are rated.

3. This Scale is an excellent predictor of rank order of all speakers when group scores are averaged.

4. The reliability of this scale is largely dependent on the ability of the individuals in the group to accurately discriminate among their observations of habits listed in each tetrad.

THE EFFECTS OF A LANGUAGE TRAINING PROGRAM ON FOREIGN SOUNDINGNESS

DALE J. LUNDEEN, PAUL H. PTACEK,
CLARK D. STARR, ERNEST H. HENRIKSON
University of Minnesota

IN recent years, the number of foreign students attending colleges and universities in the United States has increased considerably. At the University of Minnesota, the increase in foreign nationals affiliated with educational programs has been from 346 to 836 in the past nine years. As a result of this increase, it has become necessary to develop somewhat more formalized programs than those which were previously considered to be adequate in meeting the needs of the foreign student. At the present, these programs involve considerable expenditure of time, talent, and money. Therefore, an evaluation of their effectiveness has become imperative. The following is a report on one aspect of the total evaluation of the foreign language training program at the University of Minnesota.

Purpose of Study and Program of Instruction. The purpose of the present study was to determine if the language training program reduced the foreign soundingness of the speech of a group of foreign students attending the University of Minnesota. The program began with a discussion of English phonemes and how they are produced. Particular attention was centered on those which are especially difficult for most foreigners e.g. [θ], [ɪ], [r]. Ear training was given to develop discrimination between commonly confused phonemes. Commonly used words, selected from Thorndike's lists, were used for analysis and practice. Strong and weak forms of words, rules of pronunciation, and melody patterns of spoken English were explained. Practice was given in the application of the principles discussed for each of these aspects of speech. The training involved 20 hours of group work spread over a period of six months; limited individual instruction was given.

I. Procedure

A group of newly arrived foreign students was subdivided into a control group (31 subjects) and an experimental group (23 subjects). Recordings were made of the speech of all subjects before and after the six month language training program given the experimental group. The foreign soundingness of each recording was rated, and the obtained ratings were analyzed to determine the effects of the training program. An "improvement score" was determined by subtracting the final rating from the initial rating for each subject. A constant of ten was added to each improvement score to eliminate negative scores.

Population. Fifty-four newly arrived graduate students from a number of European, Middle Eastern, and Oriental countries comprised the study population. The mean age was 28 and the range was 21 to 49. The native languages used by the majority of the subjects would be classfiied as Indo-European. Their proficiency in the English language ranged from minimal basic knowledge of English to highly proficient use of the language.

Recording. The standard articulation passage used in this study was "My Grandfather," and each subject's reading of this passage was recorded on a Revere Tape Recorder. The opening section of "My Grandfather," the same words for each subject, was excerpted from each recording, and these samples were then arranged in random order on a single tape. Each sample was approximately 30 seconds in length.

Ratings. Using the aforementioned single tape, the foreign soundingness of each subject's speech was rated on a nine-point scale by six judges who were trained speech correctionists. The extremities of the scale were defined as (1) normal American speech, and (9) extremely foreign-sounding speech. The base for the latter extremity was derived by selecting speech recordings from the files of the Speech and Hearing Clinic which were judged by the staff to be extreme in foreign soundingness. These recordings were used for training the judges for their rating task. None of the recordings of the subjects being studied was used for training purposes. The reliability of the judges' ratings was determined by correlating the average rating made by three judges with the average rating made by the other three judges. A correlation of .90 was obtained.

II. Results

An analysis of the obtained data yielded the following results:

1. In the experimental group, the mean of the initial foreign soundingness ratings ($M = 4.24$, $SD = 1.24$) exceeded the mean of the final ratings ($M = 3.78$, $SD = 1.30$), and the difference between the means was significant at the 1% level ($t = 3.97$). This indicates a decrease in the foreign soundingness of the experimental group during the period covered by the language training program.

2. In the control group, the mean of the initial foreign soundingness ratings ($M = 3.47$, $SD = 1.44$) exceeded the mean of the final ratings ($M = 3.38$, $SD = 1.38$), but the difference between the means was not statistically significant ($t = 1.70$, $p = .10$). It is not unreasonable to assume that some real improvement did occur in the speech of the members of the control group, but from a statistical point of view this improvement has not been demonstrated.

3. The mean initial foreign soundingness rating for the control group ($M = 3.47$, $SD = 1.44$) was found to be less than the mean initial rating for the experimental group ($M = 4.24$, $SD = 1.24$). The difference between these means was significant at the 5% level ($t = 2.07$).

TABLE I

Analysis of Covariance of Improvement Scores (x) with Initial Foreign Soundingness Scores (y) Held Constant.*

Source of Variation	DF	Σx^2	Σy^2	Σxy	Adjusted or Reduced DF	SS	MS
Between Groups	1	64.02	278.91	133.62	1	42.29	42.29
Within Groups	52	497.46	3559.85	266.08	51	477.57	9.36
Total	53	561.48	3838.76	399.70	52		
		$F = 4.52$**					

*To simplify computations the sums of the six judges' ratings were used instead of the average of the six ratings.

**Significant at .05 level.

4. Analysis of covariance was used as a means of controlling possible variation in gains resulting from the difference between mean initial scores of the two groups (See Table I). The F-ratio resulting from this analysis was significant at the 5% level, indicating that the mean gains of the two groups were different. When the mean improvement scores of the two groups were adjusted for group differences on initial foreign soundingness ratings, the experimental group (M = 2.10) showed a greater decrease in foreign soundingness than did the control group (M = 1.79).

Within the limitations of this study, it appears reasonable to conclude that the six-month language training program at the University of Minnesota was effective in reducing the foreign soundingness of the speech of those foreign students participating in the program, and this reduction was significantly greater than that occurring in the speech of foreign students who did not participate in the program.